Motivating bureaucrats with non-monetary incentives when state capacity is weak: Evidence from large-scale field experiments in Peru

Dustan, Andrew and Maldonado, Stanislao and Hernandez-Agramonte, Juan Manuel

Vanderbilt University, Universidad del Rosario, Innovations for Poverty Action

22 December 2018

Motivating bureaucrats with non-monetary incentives when state capacity is weak: Evidence from large-scale field experiments in Peru*

Andrew Dustan
Vanderbilt University

Stanislao Maldonado†
Universidad del Rosario

Juan Manuel Hernandez-Agramonte
Innovations for Poverty Action

This version: December 22, 2018 (First version: November 6, 2017)
For the latest version of this paper, click here.

Abstract

We study how non-monetary incentives, motivated by recent advances in behavioral economics, affect civil servant performance in a context where state capacity is weak. We collaborated with a government agency in Peru to experimentally vary the content of text messages targeted to civil servants in charge of a school maintenance program. These messages incorporate behavioral insights in dimensions related to information provision, social norms, and weak forms of monitoring and auditing. We find that these messages are a very cost-effective strategy to enforce compliance with national policies among civil servants. We further study the role of social norms and the salience of social benefits in a follow-up experiment and explore the external validity of our original results by implementing a related experiment with civil servants from a different national program. The findings of these new experiments support our original results and provide additional insights regarding the context in which these incentives may work. Our results highlight the importance of carefully designed non-monetary incentives as a tool to improve civil servant performance when the state lacks institutional mechanisms to enforce compliance.

*We thank Ernesto Dal Bo, Kitt Carpenter, Claudio Ferraz, Fred Finan, Alvaro Forteza, Federico Gutierrez, Daniel Hicks, Rachid Laajaj, Mounu Prem, Santiago Saavedra, Juan Vargas and seminar participants at Universidad del Pacfico, Universidad del Rosario, Vanderbilt University and participants at the Advances with Field Experiments (AFE), Association for Public Policy Analysis and Management (APPAM), Peruvian Economic Association (PEA), Southern Economic Association (SEA), and RIDGE/LACEA-PEG conferences for insightful comments. Field implementation would not have been possible without the support of the MineduLAB and PRONIED teams at the Ministry of Education and the CUNA MAS team at the Ministry of Social Inclusion. We especially thank Jorge Mesinas, Fabiola Caceres, Luis Baiocchi, Dennis Velasco, Rosie Fontinier, Milagros Barrantes, Jorge Fernandez, Karina Tecse, Marcos Bravo, Liz Espinoza, Susan Fiestas, and Martin Blas. Juliana Aragon, Maria Paula Medina, and Cesar Huaroto provided excellent research assistance.

†Dustan: Vanderbilt University, Department of Economics, andrew.dustan@vanderbilt.edu; Maldonado: Universidad del Rosario, Faculty of Economics, stanislao.maldonado@urosario.edu.co; Hernandez-Agramonte: Innovations for Poverty Action, jmhernandez@poverty-action.org.

1 Introduction

The limited ability of the state to enforce the law and provide public goods is one of the fundamental problems facing the developing world (Besley and Persson, 2011). Whereas the roots of state incapacity are diverse,[1] there is a growing consensus that the state's effectiveness depends on the quality of those it hires to perform government functions (Bertrand et al., 2017 and Best et al., 2017).[2] But it is precisely where this quality is needed the most that civil servants are perceived to be lazy and corrupt,[3] implying that governments face fundamental challenges in recruiting and motivating civil servants to perform well. Although there is growing interest in issues pertaining to selection of public officials in developing countries (Dal Bo et al., 2013 and Ashraf et al., 2015), very little is known about effective ways to attract stronger candidates and motivate those already working in the public sector. One strategy explored by scholars is offering incentives as a tool to attract a better pool of applicants and to induce better performance, but most existing scholarship has focused on the use of monetary incentives.[4] Although this strategy has been shown to be effective, large-scale implementation is hindered by its inherent fiscal costs and the complexity of designing the right incentive structure in contexts where state capacity is low and institutional rigidities severely limit its adaptability. Moreover, multitasking in government functions limits the ability of such monetary rewards to be effective (Dixit, 2002) and the risk of crowding-out of intrinsic motivation may be important (Banuri and Keefer, 2016).[5]

Non-monetary incentives may offer a flexible and affordable strategy to increase bureaucrats' performance at a large scale, yet evidence about their effectiveness in this role is limited. This is surprising given that they have been used widely in the public sector, although few of them have been rigorously tested. Some examples include reward schemes for high achievers using social recognition (Ashraf et al., 2014) or postings (Khan et al., 2018). Although these are promising ways to motivate bureaucrats, little is known about alternative and cost-effective ways to achieve the same goal. In particular, there has been limited use of insights from behavioral economics to design interventions that induce effort by civil servants and increase their compliance with public policies.[6]

[1] After the seminal contribution of Besley and Persson (2009) and Besley and Persson (2010), a large literature has been developed on the determinants and origins of state capacity in economics. For a review of this recent scholarship, see Savoia and Sen (2014).

[2] See, for instance, Cingolani et al. (2015) for a discussion of the literature that evaluates the role of bureaucracies in state capacity.

[3] Chaudhury et al. (2006) provide evidence on the poor performance of front-line public good providers. For instance, absenteeism of teachers (19%) and health providers (35%) in developing countries were found to be very high.

[4] Dal Bo et al. (2013), Ashraf et al. (2015) and Deserranno (2016) are some of the early contributions to this literature in the case of developing countries. See Finan et al. (2017) for an overview. Evidence for the case of developed countries is extensive and results are mixed. See, for instance, Burgess et al. (2017) for the UK case.

[5] For an overview of the issues surrounding the use of incentives in the public sector, see Besley and Ghatak (2014) and Burgess and Ratto (2003).

[6] The use of behavioral insights by governments to affect economic and social outcomes is still limited but increasing, especially in the developed world. Examples include the Behavioral Insights Team in United Kingdom and the Social and Behavioral Sciences Team at the White House in United States. Some developing countries are starting to

The use of behavioral insights is particularly relevant, given increasing evidence suggesting that civil servants are affected by the same set of behavioral biases that have been documented by behavioral economists in other populations.[7] Civil servants exhibit biases in information processing, value assessment, and decision-making (Banuri et al., 2017), despite special provisions that exist in legislation that regulate the behavior of civil servants regarding qualities such as objectivity and impartiality that they are expected to follow.[8] Despite this, little has been done to incorporate these behavioral dimensions in efforts to improve civil servants' performance.[9] This is especially true for the case of large-scale interventions.

In this study, we report the results of a set of large-scale field experiments designed to analyze the effect of non-monetary incentives on bureaucrats' policy compliance.[10] These incentives, motivated by recent insights from behavioral economics, are delivered using text messages as a cost-effective tool. Specifically, we partnered with the Ministry of Education of Peru (MINEDU) to implement a behavioral-based short messaging service (SMS) campaign to induce bureaucrats from a school maintenance program to comply with the policies designed by MINEDU in this regard. Bureaucrats in this program execute a set of tasks that involves the use of monetary transfers from the central government to perform investments in school infrastructure. This is one of the most typical tasks that civil servants do: using monetary resources and converting them into a public good or service. This is also a common scenario in which there is room for corruption. Before the implementation of this behavioral-based SMS campaign, MINEDU was unable to verify the use of school maintenance funds for a large proportion of its civil servants in charge of maintenance activities. Many more

consider these issues more seriously in their policy-making. However, little has been done in addressing civil servants' potential behavioral biases. A recent report by the Behavioral Insights Team (Hallsworth et al., 2018) provides one of the first systematic attempts to address these issues.

[7]Early contributions from the field of psychology have also documented the role of these biases in the case of experts, perhaps the closest category to civil servants that have received attention in the literature. See, among others, Englich et al. (2006), Langfeldt (2004), and Stewart and Stasser (1995).

[8]For instance, the UK Constitutional Reform of 2010 established that civil servants should be guided by values such as "integrity, honesty, objectivity, and impartiality" and that their decisions should be based on "rigorous analysis of the evidence". In the US, Executive Order 12674 establishes that civil servants "...shall act impartially and not give preferential treatment to any private organization or individual" while Public Law 96-303 requires that government employees should "... never discriminate unfairly by the dispensing of special favors or privileges to anyone, whether for remuneration or not; and never accept, for himself or herself or for family members, favors or benefits under circumstances which might be construed by reasonable persons as influencing the performance of governmental duties." International organizations like the World Bank and the United Nations also have similar principles. See Banuri et al. (2017) for more details.

[9]After the publication of Thaler and Sunstein (2008), the use of nudges by governments has become widespread, but its use on public officials has remained limited (see Halpern (2016) for a perspective from the experience of the Behavioral Insights Team in the UK). One of the first attempts to recognize the importance of addressing behavioral biases with civil servants was documented in the 2015's World Development Report in the context of development professionals (World Bank (2015)). Recently, the Behavioral Insights Team produced a report analyzing the role of behavioral biases in the project delivery of the Department for Transport in the UK (Behavioral Insights Team (2017)). Similar efforts for the case of developing countries are scarce.

[10]We use the term "incentives" in a general way as "... something, often money or a prize, offered to make someone behave in a particular way" (Cambridge Business English Dictionary). Interestingly, despite the pervasive use of the concept in the modern economic literature, it is relatively new and somewhat problematic. See McCaffrey (2014) for a discussion.

reported the use of maintenance funds late or in an incomplete manner.[11] Moreover, a sizable group of maintenance civil servants did not even withdraw the transfers from accounts at the National Bank designed for this purpose (15%), representing a proportion of 10% of the total budget assigned for maintenance activities. Inducing compliance in this setting is costly. Upstream government bodies at MINEDU have limited information about the performance of maintenance civil servants and punishing misuse of funds is expensive since any legal action against corrupt officials would cost more that the actual funds transferred. When civil servants operate in a poor institutional environment, the use of standard means to induce compliance is not feasible due to weak state capacity.

Peru offers an ideal setting to explore the impact of behavioral-based SMS campaigns to enforce civil servant compliance. First, the country has a long history of weak governance and corruption. Civil servants are perceived as particularly corrupt by citizens, as suggested by household surveys. For instance, 26% of Peruvian citizens reported having paid a bribe.[12] It is also a country where little is done to control corruption, according to the World Bank's governance indicators,[13] and where citizens do not trust government agencies.[14] A second reason is that compliance problems are important. In the two national programs we collaborated with, around 30% of civil servants failed to comply with relevant tasks related to the provision of public goods for citizens. This is particularly critical in the context of our experiments where citizens' needs are acute. For instance, civil servants at MINEDU failed to invest school maintenance funds in a context where the educational infrastructure gap is estimated to be US$ 34.4 billion.[15] Third, MINEDU has good administrative records with information on program activities and maintenance transfers to schools. This facilitates a large-scale intervention and allows for a cost-effective and high-frequency data collection process. The final reason is the existence of a research unit at MINEDU, known as MineduLAB, that was fundamental to the implementation of the innovations tested in this paper. MineduLAB is one of the few innovation labs for educational policy that exists in the developing world.

The large-scale experiments are designed to induce compliance among civil servants using behavioral insights. In the first large-scale experiment (which we call the "Benchmark Experiment"), text messages are crafted in a way that incorporates behavioral insights in dimensions related to

[11] Before the intervention, 11% of the civil servants (more than 6,000) failed to report the use of the transfers and 9% of them declared its usage late. This meant that about 12,000 civil servants failed to comply with the existing regulations regarding the reporting of the expenses in school maintenance.

[12] Authors' calculation based on the National Household Survey (ENAHO) 2017. This calculation was performed using the number of respondents that reported having interacted with a civil servant, regardless the state agency, during 2017.

[13] In 2015, Peru ranked 35 in the percentile-rank scale of the "control of corruption" indicator, a poor performance with respect to other high-middle income countries. For instance, OECD countries ranked 85 in the same scale. See `http://www.govindicators.org` for details.

[14] According to the 2017 ENAHO survey, 57% of the surveyed citizens report that they do not trust MINEDU. The levels of distrust of other branches of the government are similar or worse. For instance, 78% of citizens do not trust the National Police and 79% do not trust their local government.

[15] A recent World Bank report estimated the educational infrastructure gap of PEN 113.5 billion. See World Bank (2017) for details.

information provision, social norms, and weak forms of monitoring and auditing. Maintenance civil servants across the country are exposed to one of five multi-message campaigns highlighting one of these dimensions. In the first treatment group, civil servants receive alerts and a link to obtain additional information about maintenance activities. The second group is informed about the amount of the transfer available in their accounts at the National Bank. Civil servants in the third group receive a message with a social norm regarding the level of compliance of other civil servants in their reference group. In the fourth group, civil servants are informed that their names will appear in a public list if they fail to comply with the rules governing the maintenance activities. Finally, the last group receives a message indicating that they may eventually be audited.

We find that these messages are a very cost-effective strategy to enforce compliance among civil servants. Receiving any message is associated with a reduction of about 20% in the compliance gap (the distance between the current levels of compliance and full compliance) for reporting expenses and 10% for withdrawal of funds from the National Bank accounts. All behavioral contents seem to be effective, although social norms and monitoring seem to be particularly effective in inducing compliance.

We then run a second experiment in 2016 (called the "Follow-Up Experiment") to further explore the role of social norms and the salience of social benefits along with other implementation details that might be relevant to transform this campaign into a public policy. Regarding social norms, we introduce the distinction between descriptive and injunctive norms (Cialdini and Goldstein, 2004), taking into account evidence suggesting that they may have differential effects (Cialdini, 2007). We also vary the reference group (school district versus the country as a whole) and the use of quantitative or qualitative norms. This approach allows us to learn more about what type of social norm may be most effective in inducing compliance.

The second change from our original design is the introduction of the dimension of social benefit as an additional treatment arm. Scholars have shown that the salience of this dimension affects selection into civil service and performance (Ashraf et al., 2015). We design text messages that emphasize the importance of having good infrastructure for students' well-being, for the pride of the school community, and for students' learning. These dimensions complement our understanding of the role of social benefit as a factor that can foster compliance among civil servants.

Two additional variations are implemented in the Follow-Up Experiment to address critical details for implementing the lessons of the Benchmark Experiment as a public policy. First, we explore the issue of learning and the potential effectiveness of a behavioral-based SMS campaign on inducing compliance over time. One concern is that civil servants can be "fooled" only once, but will update their priors regarding the weak nature of any threats emphasized in the text messages or otherwise become desensitized to any novel material introduced by the messages, making them ineffective beyond the first iteration. We find that previous exposure to the campaign during the Benchmark Experiment has no effect on outcomes in the Follow-Up Experiment, suggesting the lack of learning effects. Second, we study whether the duration of the SMS campaign matters. We vary the number of SMS delivered to civil servants, designing a short duration campaign of three

messages and a long duration one with seven messages delivered with the same frequency. We find no evidence that the short duration campaign was less effective, suggesting the existence of saturation effects for the case of the long duration treatment.

We also run a third experiment (which we call the "External Validity Experiment") with a different national program in order to explore external validity issues. Our large-scale intervention does not suffer from the standard external validity problem in which only a small subset of a population is part of an experiment. However, we want to learn whether the basic results of our Benchmark Experiment can be replicated in a population of civil servants with different characteristics. As has been recognized elsewhere,[16] public sectors in the developing world are characterized by the presence of civil servants in environments with fixed and rigid hiring and promotion practices, along with an increasing number of civil servants hired under temporary and more flexible contracts. Because of the relevance of this issue, we run a variant of the Benchmark Experiment with civil servants from the Family Supporting Services of the National Program CUNA MAS, an early childhood development intervention run by the Ministry of Social Inclusion (MIDIS). Civil servants from this program are typically subjected to less rigid short-term contracts, usually linked to less generous benefits packages than in the education sector. The program faces an issue of poor compliance in the submission of monthly reports regarding the delivery of services by the personnel in charge of interacting with front-line providers. This activity is performed by the program's field monitors and a large fraction of them fail to comply fully with the reporting of this information, affecting the ability of the program to plan service delivery for the next month. We test the role of social norms and monitoring, the most promising interventions found in the Benchmark Experiment, in a monthly intervention implemented from September 2016 to January 2017 with all the civil servants in charge of this task across the country.

We find no role for social norms in improving compliance among these civil servants, but a positive impact of monitoring. We estimate an impact of 21.4% in terms of reducing the compliance gap for the monitoring treatment. We interpret this result as evidence of the critical role of tenure in understanding the impact of these treatments. We hypothesize that, in contrast to civil servants in the education sector, CUNA MAS's officials are less sensitive to messages that emphasize a social norm because their expectations of keeping their posts in the future are low. As a consequence, they give a low weight to their peers' perceptions when deciding to comply with the program's policies. However, the monitoring treatment plays a role in this setting because officials do care about the information the program's upstream bureaucrats have regarding their performance. Therefore, the institutional characteristics of the branch of government where these civil servants work are critical to understanding the power of non-monetary incentives.

Further results from the Benchmark and Follow-Up Experiments are informative about civil

[16]The existence of an important fraction of temporary workers in the public sector is a characteristic that has persisted in the developed and developing world, despite efforts to minimize patronage and political control (See Grindle, 2012). A recent report by the UK Department for International Development, based on surveys of civil servants in Africa, Asia, Eastern Europe and Latin America, has estimated that the fraction of temporary workers is about 23%. See Meyer-Sahling et al. (2018) for details.

servants' responses to the SMS campaign and the potential cost-effectiveness of such interventions. Comparing expenditures documented in the expense reports with those planned in initial phases of the maintenance cycle, we find that exposure to the SMS campaign resulted in a small increase in the level of reported expenditures in filed reports, while the composition of expenditures between categories remained essentially unchanged. In particular, there was no shift of expenditures from hard-to-implement expenditure categories (e.g. electrical or sanitary upgrades) toward those that are fast and easy to implement, such as classroom supply purchases. This suggests that the campaign did not pressure civil servants to comply by deviating from the approved plan in favor of expedient completion of the mandated tasks. Using supplemental data from a school census, we find no systematic evidence of appreciable impacts on infrastructure quality, as would be expected given the transfer amounts. The program was cost-effective: for every $1 spent on the program (including estimated labor costs of administering the program), about $800 were reported to PRONIED in a timely manner.

The results of these three experiments provide critical insights regarding the use of behavioral-based SMS campaigns to enforce policy compliance in the public sector. In fact, MINEDU has used the insights of these experiments in designing a national policy to motivate civil servants to comply with PRONIED's policies. CUNA MAS is also exploring steps to institutionalize the results of the External Validity Experiment in policies to motivate its officials on a regular basis.

This paper relates to an emerging literature about the personnel economics of the state (Finan et al., 2017) and bureaucracies in developing countries (Pepinsky et al., 2017).[17] In particular, our paper is related to a growing literature about strategies to increase civil servant performance. Existing contributions have emphasized the role of monetary incentives in the case of enforcement agents (like tax collectors as in Khan et al., 2016) and front-line service providers like teachers and health professionals. A large body of evidence has been collected for the latter case regarding incentives based on outcomes (like test scores and health measures) as well as inputs (such as attendance and service delivery), with mixed results (Hasnain et al., 2014 and Finan et al., 2017).

Evidence on the role of non-financial incentives for civil servants is scarce. Some scholars have explored the role of transfers (Banerjee et al., 2014, for the case of the police force in India), promotions (Karachiwalla and Park, 2017, for the case of teachers in China) and postings (Khan et al., 2018, for property tax inspectors in Pakistan) with positive results. More in line with the approach followed in this paper, others have used non-financial rewards such as in-kind prizes (Glewwe et al., 2010) and social recognition (Ashraf et al., 2014) with promising results. But evidence on this subject remains limited, leading Finan et al. (2017) to remark in their review that there is the need to explore it further. This paper contributes to fill this gap.

This paper is also related to a very new literature that explores the role of behavioral biases in the case of civil servants. Recent papers have systematically documented the existence of present bias (Andreoni et al., 2016), status quo bias (Celhay et al., 2015), confirmation bias, framing

[17] Recent reviews in the political economy literature include Dal Bo and Finan (2016) and Azulai et al. (2014).

effects, inattention bias and optimism bias (Hallsworth et al., 2018 and Banuri et al., 2017) among bureaucrats. Rather than documenting the existence of such biases, this paper tests a set of strategies to deal with them.

This paper is also related to a literature regarding the use of digital technology to improve policy outcomes. Researchers have used smartphones and other technological devices to monitor bureaucrats' attendance (Callen et al., 2018, Dhaliwal and Hanna, 2017, and Duflo et al., 2012), increase accountability (Aker et al., 2017), improve public service delivery (e.g. Dal Bo et al., 2018) and minimize corruption (Muralidharan et al., 2016 and Lewis-Faupel et al., 2016). This paper contributes an innovative way to use text messaging to induce compliance among civil servants.

Finally, this paper is related to a new but growing literature about experimentation at scale (Muralidharan and Niehaus, 2017 and Davis et al., 2017). There are several reasons why large-scale experiments are important, among them the search for external validity.[18] A very recent literature has emphasized the issue of generalizability of the results from small-scale experiments to a relevant population using theoretical approaches (e.g. Banerjee et al., 2017) or statistical designs based on reweighting (e.g. Andrews and Oster, 2017) or bounds (e.g. Kowalski, 2018). This paper's design addresses several of these concerns regarding external validity. By implementing the interventions with a population of civil servants, concerns about scalability and generalizability of results from small samples are not present. By implementing the Follow-Up Experiment one year later than the original intervention, it also addresses the concern that its results are not externally valid due to time-specific aggregate shocks (Rosenzweig and Udry, 2016). Furthermore, by implementing a version of the Benchmark Experiment with a different population of civil servants, it also addresses the issue of how generalizable the results are for alternative settings and populations. We acknowledge that our design cannot fully address other dimensions of external validity, but in this respect it is at the frontier of the experimental literature in development economics and other fields (Peters et al., 2018).

The rest of the paper is organized as follows. Section 2 provides some basic details about the institutional setting. Section 3 introduces the research design and Section 4 presents the estimation strategy. Section 5 describes the results of the Benchmark Experiment and Section 6 presents the results of the Follow-Up experiment. Section 7 gives the results of the External Validity Experiment. Section 8 presents additional results (including a cost-effectiveness analysis) and Section 9 concludes.

[18] Scholars have explored several dimensions of external validity, in particular the scalability of interventions, the existence of market equilibrium effects and externalities, site selection and piloting bias, the effect of treatments on different populations, the effect of treatment in the same population under different circumstances, and the effect of different, but related, technologies (Allcott, 2015; Banerjee et al., 2017; Banerjee et al., 2017; and Al-Ubaydli et al., 2017). See Peters et al. (2018) for an overview of these issues in the development literature.

2 Institutional setting

The first field experiment was run in collaboration with the National Program of Educational Infrastructure (PRONIED) at the Ministry of Education (MINEDU) in Peru. The goal of this program is to expand, improve, replace, rehabilitate and construct public educational infrastructure, including both buildings and furniture. One of its critical functions is the School Infrastructure and Furniture Maintenance Program, which entails the allocation of monetary transfers to a civil servant in charge of regular maintenance activities in each school. The cycle of the intervention is summarized in Figure 1.

The program has a strong participatory component. At the beginning of each school year, a Maintenance Committee is created with the participation of teachers, students and parents. This committee is the unit at the school level in charge of implementing all activities related to the maintenance of basic infrastructure and furnishings. In addition, an Oversight Committee is also formed along the same lines. A coordinator of the Maintenance Committee is chosen among the teachers (including the principal).

Once this institutional structure is in place, a technical form is produced with the maintenance investments that are considered relevant for the Maintenance Committee. This technical form is submitted to the Educational Local Management Unit (UGEL) for approval. An UGEL is similar to a school district in the US case. Once the technical form is approved, maintenance civil servants are able to start carrying out maintenance activities and withdraw maintenance funds from accounts at the National Bank assigned to them exclusively for investments in infrastructure. Once the execution phase is completed, they are required to write an expense report accompanied by invoices for all expenditures, along with a report prepared by the oversight committee, which evaluates whether they believe that funds were used for their intended purpose. Unused funds are returned. The final step is the approval of the expense report, an activity performed by an infrastructure specialist at the corresponding UGEL.

It is important to emphasize that the maintenance funds are assigned to a single maintenance civil servant in each school. While Maintenance Committees are composed of teachers, students and parents, only a single maintenance civil servant (typically the principal but sometimes a teacher) is formally assigned the role of receiving the maintenance funds via an account at the National Bank. For this reason, the number of maintenance civil servants is the same as the number of schools that were part of these experiments.

3 Research design

We partnered with MineduLAB, an innovation lab that exists inside of MINEDU, to design a cost-effective strategy to address civil servants' non-compliance with PRONIED's rules.[19] We

[19]MineduLAB is an innovation lab for education policy in Peru, created by Innovations for Poverty Action (IPA), the Abdul Latif Jameel Poverty Action Lab (J-PAL) and the Ministry of Education of Peru. For more details, see

implement this strategy in a large-scale field experiment with the population of civil servants in charge of the school maintenance program for which a cellphone number is recorded in MINEDU administrative records. Based on conversations with the MineduLAB and PRONIED teams, we focused on two critical variables: withdrawal of maintenance transfers and reporting of expenses. As discussed above, these two variables are the most relevant in the context of this intervention. In particular, these two outcomes are those for which the maintenance civil servant is most directly responsible (as opposed to oversight committees or UGEL officials) and correspond most closely to observable actions toward carrying out the required maintenance (in the case of withdrawals) and accounting for expenses incurred in performing the maintenance (in the case of submitting the expense report). We will consider additional outcome variables that are also available in the administrative data, but these two will receive special care.

3.1 Treatment

We implemented a SMS campaign to increase civil servants' compliance with maintenance activities. In 2015, we performed the Benchmark Experiment in which we tested a set of basic behavioral insights. We describe the details of this experiment in this section.

Each SMS contains a fixed and a variable component. The fixed component includes the bureaucrat's first name and the deadline for task compliance. These fixed elements are rooted in behavioral insights. The use of personalized messages has been shown to be an effective strategy (Karlan et al., 2012). On the other hand, the use of exogenous deadlines has been proven to be more useful when agents suffer from procrastination (Ariely and Wertenbroch, 2002). The variable component is the main behavioral lever that we use to induce a change in bureaucrats' behavior. We describe this component below.

Maintenance bureaucrats are assigned to one of six groups. Bureaucrats in the control group receive no SMS. The remaining bureaucrats receive an SMS with behavioral content at fixed points during the intervention cycle. In total, each bureaucrat in any of the treatment groups receives up to five SMS. These SMS share the same behavioral insight over the cycle but vary in terms of the type of maintenance activity that is emphasized. For instance, near the beginning of the intervention cycle, bureaucrats receive SMS emphasizing the withdrawal of maintenance funds, whereas near the end of the cycle, SMS emphasize the filing of expense reports. Bureaucrats only receive a particular message if they have not complied with the activity being emphasized in that SMS. The full set of messages is presented in Table S2 of the Online Appendix.

Bureaucrats in the *reminder/warning treatment* receive SMS with an alert and the URL of the PRONIED website where the bureaucrat can obtain more information. Reminders are one of the most popular tools used in behavioral science to influence behavior and the inclusion of an alert is motivated by the need to prime a sense of urgency to comply with maintenance activities. Reminders are motivated by the existence of limited attention problems and are tools that can

`https://www.poverty-action.org/minedulab`.

potentially change the inter-temporal allocation of mental resources to enforce compliance.[20]

Bureaucrats in the *monitoring treatment* receive SMS with information regarding the amount of transfers not yet withdrawn from the bank or not yet declared on the expense report, depending on the timing of the message in the intervention cycle. This information gives bureaucrats the impression that their actions are being observed and, as a consequence, may induce them to comply with maintenance policies. This treatment should not be surprising for a fully rational agent since it is perfect knowledge that the program is able to observe funds withdrawal and expense reporting. Therefore, by making salient a fact that is common knowledge among civil servants, it is possible to re-create some critical dimensions of monitoring systems in a cost-effective way.[21]

Bureaucrats in the *social norm treatment* receive SMS with a message emphasizing that most bureaucrats are complying in their reference group (UGEL). Social norms are understood in this paper as a set of informal rules and unwritten codes that establish what we expect of others and what others expect from us (Young, 2015).[22] Following Cialdini and Goldstein (2004), it is possible to establish a useful distinction between norms that inform us about what is typically done (descriptive norms) and norms that inform us about what is typically approved or disapproved (injunctive norms). We used a qualitative descriptive norm to minimize the risk of backlash effects, considering a body of evidence that suggests that providing actual levels of conformity with a social norm can induce more people to deviate from it if their baseline expectations regarding conformity with the norm were higher.[23] In the follow-up experiment we further explore variants of social norms, including quantitative norms and alternative reference groups.[24]

Bureaucrats in the *shaming treatment* receive SMS with information regarding the possible publication of a list with the names of those bureaucrats who fail to comply with the reporting of expenses. The goal is to induce concern regarding potential reputational loss in order to motivate compliance (Eyal, 2014), especially when baseline non-compliance behavior is deeply rooted. This treatment arm is based on a large body of evidence indicating that people are more likely to

[20]Karlan et al. (2016) propose a theoretical model to justify the use of reminders in the context of saving based on the idea that individuals misunderstand the value of future consumption and then under-save or under-borrow. From an empirical point of view, reminders have been proven to be useful in inducing donations (Sonntag and Zizzo, 2015), take-up of social benefits (Bhargava and Manoli, 2015), gym attendance (Calzolari and Nardotto, 2017), electricity consumption (Alcott and Rogers, 2014), savings (Karlan et al., 2016), and adherence to medical treatments (Vervloet et al., 2012).

[21]It has been shown that monitoring is an effective strategy for improving performance (Callen et al., 2014) and controlling corruption (Olken, 2007), but it is costly and can be captured by corrupt officials (Finan et al., 2017). There is a large literature about the use of monitoring mechanisms by the government and citizens. See Molina et al. (2017) for a review.

[22]There is a large literature on social norms in economics, sociology, psychology, philosophy, legal studies, political science and anthropology. Given the behavioral approach used in this paper, a psychological approach is emphasized. For an overview of this literature, see Mackie et al. (2015) and the references therein.

[23]For a discussion about backlash effects, see Miller and Prentice (2016)

[24]Social norms have been proven to be effective in inducing behavioral change in domains such as recycling (e.g. Schultz, 1999; but also see Chong et al., 2013), energy consumption (e.g. Alcott and Rogers, 2014), water use (e.g. Ferraro and Miranda, 2013), smoking and drinking (Foxcroft et al., 2015; Hansen and Graham, 1991), sexual practices (e.g. Lynch et al., 2004), domestic violence (e.g. WHO, 2009), female labor supply (e.g. Bursztyn et al., 2018), voting (e.g. Gerber and Rogers, 2009), charitable giving (e.g. Frey and Meier, 2004), and tax compliance (e.g. Hallsworth et al., 2017).

comply when their behaviors are observed (Rogers et al., 2018). This insight has been used in health interventions to induce behavioral change of non-healthy behaviors such as open defecation (Gertler et al., 2015) and smoking (Voigt, 2013), but it has been used in a variety of settings.[25]

Finally, bureaucrats in the *auditing treatment* receive SMS with a soft threat of auditing. Specifically, they are told that they will be visited for supervision of their maintenance activities. Schools are already visited on a regular basis by UGEL representatives for several matters, including (of course) the development of maintenance activities. In that sense, the intervention is simply making salient an event that civil servants will face over the course of the academic year. However, given the scale of the intervention, the probability of facing a visit is low at a given moment of time. We take advantage of this fact to induce compliance among civil servants by reminding them about the fact that they will be visited by UGEL officials.[26]

Figure 2 provides an example of the content of the SMS messages. Table S2 and Figure S2 in the Online Appendix present, respectively, the detailed contents of all SMS delivered over the intervention cycle in 2015 and the details about critical dates during the SMS campaign.

3.2 Data

This section discusses the data sources and variables used in this experiment. We combine different administrative records for implementing the research design and to evaluate the impact of the intervention. These data are complemented with surveys that are typically carried out by MINEDU for other purposes. As mentioned above, we consider information for all schools with a maintenance civil servant for which a cellphone number is available in MINEDU's administrative records. Figure 3 presents a map with the location of these schools across the country.

3.2.1 Data sources

To implement and evaluate the quality of our randomization strategy, we exploited the School Census that is carried out annually by the Educational Statistics Unit at MINEDU. This census collects information from all public and private schools in the country and includes information on enrollment, students' performance (progress, promotion, repetition, drop-out rates, etc.), teacher and school characteristics. We used the 2014 Census in the design.

Information on outcomes was obtained from two main administrative data sources. To monitor compliance with the maintenance activities, we used the WASICHAY system, which is PRONIED's school maintenance management system. This system was designed for maintenance civil servants to record and update information related with all maintenance activities, including the uploading

[25] Among other topics, scholars have explored the role of shame on voting (e.g. Gerber et al., 2010), environmentally friendly behaviors (e.g. Delmas and Lessem, 2014), and charitable giving (e.g. Karlan and McConnell, 2014).

[26] There is a growing body of evidence about the effectiveness of audits. Most papers in this literature have explored the role of audits in making politicians accountable (e.g. Ferraz and Finan, 2008) and reducing political corruption (e.g. Bobonis et al., 2016). Audits have been also used to induce legal compliance by firms (e.g. Duflo et al., 2013) and to induce better public service provision among local politicians (De La O and Martel, 2015). We are not aware of studies using soft forms of auditing with civil servants as used in this paper.

of maintenance technical files and expense reports. To analyze the impact of our intervention on PRONIED funds withdrawal, we had access to balance information for all of the accounts created by PRONIED on behalf of the maintenance civil servants at the National Bank of Peru. A great advantage of using these administrative records is the access to detailed information during different parts of the process and the minimization of attrition problems that are typical of many field experiments.

We complement these data with the SEMAFORO survey. This is a rolling census that covers all public schools during the academic year to collect information about the provision of educational services. Each month, a random group of schools is visited by MINEDU monitors until all public schools are eventually covered at the end of the academic year. We take advantage of the random component in the selection of schools to be surveyed over the academic year to compare schools in the control and treatment groups to assess whether treatment affected the quality of infrastructure, one of the components evaluated in the SEMAFORO survey.

3.2.2 Main variables

The main variables used in this study are constructed based on compliance with the infrastructure maintenance policies. Table S1 in the Online Appendix contains a full list of all the variable definitions used in the Benchmark Experiment. We construct a set of dummy variables for compliance with each step of the maintenance cycle. The most important variable is a dummy for whether the maintenance civil servant submitted the expense report. This is the one for which maintenance civil servants are accountable. We also consider completion of the oversight report and the approval of the expense report. These two later outcomes are not directly under the control of the maintenance civil servants, but they provide some measure of the quality of their performance since it is expected that these reports are more likely to be approved when the maintenance activities are performed correctly. This is an imperfect measure of quality, however, since imperfect compliance by UGEL monitors in reviewing these reports on time can affect their approval.[27] Because the final step that the maintenance civil servant carries out is the submission of the expense report, we limited the experimental sample to the universe of civil servants who had not already completed this step at the time that the SMS campaign began.

We also create dummy variables for different levels of compliance with the withdrawal of maintenance funds at the National Bank. We consider the withdrawal of any positive amount as well as withdrawal of at least 50%, 80%, 90% and 95% of the transferred funds.

We use these administrative records to estimate the impact of the SMS campaign for different periods of time. We present estimates for different points in times during the SMS campaign as well as for time periods after the campaign was completed.

[27] Monitors at the UGEL level are in charge of evaluating a large number of expense reports. This introduces delays in completing the evaluation. Therefore, whether a report is approved by the time that we are considering in the analysis is partly due to these delays and not only caused by the inability of the maintenance civil servant to produce reports of sufficient quality.

3.2.3 Summary statistics

Table 1 reports summary statistics for the benchmark experiment. Panel A presents information about pre-treatment outcomes. At week 20 of the PRONIED project cycle (with the end of week 30 representing the end of the cycle and the deadline for filing expense reports), a large proportion of bureaucrats have already submitted the forms necessary to receive their transfers and begin maintenance activities. For example, 86% have gone through the relevant steps with the oversight committee and 67% have filed their commitment act, which signals that they intend to carry out an approved plan of work. By construction, nobody in this sample has submitted an expense report. Before the start of the intervention, National Bank balances were on average PEN 2,700 (close to US$ 820).

Panel B presents outcomes at week 30, the official close of the project cycle. As expected, compliance outcomes are better. More than 76% comply with the required submission of the expense report. Somewhat higher compliance levels are found for submission of the commitment act (84%). Outcomes under the control of UGEL officials show relatively low levels of compliance, as expected.

Panel C presents information about bureaucrats' characteristics. 45% of the maintenance civil servants are male with an average age of 46 years. Close to 30% of them are appointed civil servants and receive about PEN 7,700 (more than US$ 2,100) as funds for maintenance activities.

Finally, Panels D and E present school and district level characteristics. The average number of classrooms is 5 (with a standard deviation of 6.2), which suggests that schools in Peru are typically small. This is consistent with the average of 128 students. Infrastructure quality is relatively poor: most schools do not have bathrooms connected to a public drainage system, while leaks and water infiltration are common. These schools are located in districts that are mostly rural, although a high proportion of them are connected to electric service. Access to internet and bank branches in the district is also low on average.

3.3 Randomization

Assignment to treatment was randomized at the school level. To implement this design, we exploited school census data and other administrative records to evaluate randomization balance. We proposed a simple randomized design to PRONIED for two reasons. On one hand, the sample size of our experiment limits the potential gains from more elaborate randomization methods. As discussed in Bruhn and McKenzie (2009), all randomization methods for sample sizes higher than 300 units deliver very similar results. With more than 24 thousand schools in our benchmark experiment, we are clearly beyond this threshold.[28]

The second reason is the role of spillovers. We do not expect spillovers to be an issue in this

[28] In the Follow-Up and External Validity experiments we considered block-randomized designs to take into account the fact that a large number of treatment conditions are tested and a lower number of civil servants is available, respectively.

setting. Bureaucrats from MINEDU receive SMS on a regular basis for various matters unrelated to this experiment. We do not expect them to share their SMS with their colleagues in other treatment arms, given that receiving these messages is not notable in itself.[29] Even if we were concerned about this issue, there is little reason to believe that a cluster-randomized design would have successfully mitigaged this problem. Cluster designs are useful for spillovers that depend on physical distance, which in this setting may not be the relevant dimension. Given the characteristics of the bureaucracy at MINEDU, with long-term horizons and with regular rotations across the same UGEL or region, there is a high chance of some level of contamination. Even if that is the case, any positive result can be interpreted as a lower bound on the treatment effect.[30]

Table 2 reports the means and standard deviations for pre-treatment characteristics by treatment status. We consider a large set of variables and find that all treatment groups are balanced on pre-treatment characteristics, with two exceptions. We find evidence of imbalance at the 10% level in terms of the proportion submitting the commitment act by week 20, although the magnitude of this potential imbalance is very small, and in district altitude. We show that results are robust after controlling for exogenous covariates.

4 Estimation

To evaluate the effect of the different messages on civil servant compliance, we estimate:

$$y_{smu} = \alpha + \sum_{j=1}^{5} \beta_j treat^j_{smu} + X_{smu}\delta + \varepsilon_{smu} \tag{1}$$

where y_{smu} measures, for school-maintenance civil servant s in municipality m and UGEL u, the outcomes of interest in terms of compliance at different stages of the maintenance cycle as well as the withdrawal of funds from the National Bank accounts. The term $treat^j_s$ denotes the five treatment groups and X_{sm} is a vector of school/civil servant s and municipality m characteristics that may be correlated with compliance. These characteristics include location, number of classrooms, number of students, allocated funds, gender, age, and other school and municipality-level characteristics. Standard errors are clustered at the UGEL level.[31] The coefficients β_j recover the causal effects of the SMS treatments on compliance outcomes as long as treatments are orthogonal to ε_{smu}. Despite randomization of treatment, this assumption might fail due to non-compliance, attrition,

[29] At the time we implemented these experiments, civil servants from MINEDU were regularly exposed to the use of SMS as part of MINEDU's communication policy.

[30] Moreover, a recent theoretical paper by Savje et al. (2017) concludes that, for scenarios with limited or even moderate spillovers, standard estimators are "good enough" to recover causal effects as long as the sample size is very large, a condition fulfilled in our experiments. Savje et al. (2017) propose a new parameter robust to the presence of spillover effects and show that standard estimators for the average treatment effect converge to this new parameter when the sample size converges to infinity. See Savje et al. (2017) for details.

[31] The standard practice consists of clustering standard errors at the treatment level (school in this case). We follow a more conservative approach of clustering the standard errors at a higher level because some of the study outcomes are determined by officials at the UGEL level.

or spillover effects. We discuss them below.

Several factors can compromise the validity of our research design. First of all, civil servants might fail to comply with their treatment assignment. This is very unlikely given the nature of our interventions, which do not represent a particular status that can be rejected by the subjects of our field experiment. They can behave in response to the message contents assigned to them but they cannot receive a different content by choice. Therefore, there is no room for the traditional compliance problems or the use of common solutions like local average treatment effect estimators as in Angrist et al. (1996).

Secondly, attrition problems do not pose a problem, given the use of administrative records to evaluate impacts. We are able to follow compliance with the maintenance activities for each of the maintenance civil servants in real time using the WASICHAY system. We are not able to observe a very limited number of maintenance civil servants for which information about their National Bank account was not recoverable for a variety of reasons.[32] In the Online Appendix, we account for attrition using the bounding method developed by Lee (2009).

Finally, spillovers may play a minor role in our setting. As described in section 3.3, we do not expect spillovers to be critical given the fact that civil servants from MINEDU receive SMS on a regular basis. Widespread sharing of SMS among them is unlikely. Even so, our results would be interpreted as lower bounds of the true treatment effect.

5 Effects of non-monetary incentives on policy compliance

We start by estimating equation 1 for policy compliance outcomes and the National Bank balance data. We then provide evidence on the dynamic behavior of our treatment effects, taking advantage of frequent data from administrative records for the intervention and post-intervention periods.

5.1 WASICHAY outcomes

Table 3 presents the results for the compliance outcomes. Column 1 gives the effect of the SMS on the submission of the expense report, omitting municipality, school and bureaucrat covariates. Reported coefficients correspond to percentage point changes (that is, the dummy dependent variables are multipled by 100). Receiving a message, no matter its content, increases by 3.86 percentage points the probability of submitting the expense report by the deadline.[33] This result is strongly statistically significant and represents a reduction of 20.9% in the compliance gap. Estimating instead separate treatment effects for each message type, we find that all treatments

[32] We were not able to recover information from the National Bank accounts for 3,245 maintenance civil servants in our Benchmark Experiment in 2015. This was not an issue in the 2016 intervention, in which only one civil servant was not found in the National Bank database.

[33] By providing estimates for the average SMS along with those of specific treatments, this specification also addresses concerns of multiple testing.

are strongly significant, with coefficients of similar magnitudes. This suggests that civil servants responded to similar degrees to the different behavioral principles that were used to design the SMS campaign.

Columns 2 and 3 present results for submission of the oversight report and the approved expense report, again omitting covariates. As discussed above, the approval of these reports is performed by UGEL officials who are not directly exposed to our non-monetary incentives. We find that in both cases, UGEL officials are more likely to have approved an expense report, as suggested by the statistically significant coefficients for receiving any message. This indicates that treated civil servants not only comply with the policy, but did so in a way that increases the (unconditional) probability of having a report approved by their UGEL. We also find that the reminder/warning and the social norm treatment were the ones that induced this response. Descriptively, the role of social norms appears large, and when also considering its large estimated effect on expense report submission, social norms seem to be quite effective in inducing civil servants to comply with the maintenance policies.

Columns 4 to 6 report the same outcomes as Columns 1 to 3 but adding controls to the basic specification. These controls include pre-treatment bureaucrat and school characteristics as previously described in detail in section 3.2.3. Standard errors decline somewhat, as expected, and point estimates are similar but are slightly larger in most cases.

5.2 National Bank balances

Table 4 presents results on the withdrawal of funds from National Bank accounts assigned to each maintenance civil servant. This allows us to analyze whether the intervention induces maintenance civil servants to withdraw monetary transfers to be invested in school infrastructure maintenance.

Column 1 presents the results for whether maintenance civil servants withdraw any positive amount of funds, from a specification excluding covariates. We find no effect of our interventions on withdrawal of funds. This is not unexpected since almost all civil servants withdraw at least something (the control mean is 99.7%). A similar result in found for withdrawing at least 50% of the transferred funds (Column 2). Columns 3 through 5 show a positive effect of the SMS campaign on the withdrawal of funds. The campaign caused a 0.92, 1.05 and 1.46 percentage point increase for withdrawal of at least 80%, 95% and 99%, respectively. These effects represent a reduction in the compliance gap of 10%, 10% and 13% with respect to the control group, respectively.

These results are driven by the reminder/warning, monitoring and shaming treatments. The monitoring treatment has the largest point estimates (18.9% in terms of reducing the compliance gap for withdrawing at least 99% of funds). Point estimates for the reminder and shaming treatment are of slightly smaller magnitude, but we are unable to distinguish statistically between the effects of any of the treatments.

Columns 6 to 10 present results for the specification with controls. Again, point estimates are

slightly larger and standard errors smaller. Qualitative results are essentially the same, with the exception that the reminder treatment is now also statistically significant at the 10% for withdrawing at least 95% of funds. As previously discussed, we were not able to recover data on these outcomes for 3,245 maintenance civil servants. To verify that this source of attrition is not driving our results, we implement the bounding strategy proposed by Lee (2009). Results are reported in Table S9 in the Online Appendix. Our results are robust to this form of attrition.

5.3 Dynamics of effects

We take advantage of detailed administrative records to explore the dynamics of the SMS campaign treatment effects. Due to space constraints, we focus on the average effect of receiving a SMS regardless of its behavioral content. We also restrict the analysis to two outcomes: submission of the expense report and withdrawal of at least 99% of funds. The Online Appendix presents the results for all the other outcomes.[34]

Figure 4 shows the dynamics of treatment effects for submission of the expense report. The horizontal axis shows the date for which the treatment effect was estimated. The dashed lines indicate the SMS campaign period from August 15th to October 1st. Pre-treatment data covers the two weeks before the start of the SMS campaign and post-treatment data includes several weeks until December 31st. The vertical axis reports the effect of the SMS campaign in percentage points. Before August 15th, we observe no differences between treatment and control groups. We do observe differences after the beginning of the intervention, nearing their peak after three weeks and remaining similar until the deadline. The effect persists even several weeks after the end of the SMS campaign. The 95% confidence intervals do not contain zero until December. There are two reasons for the declining treatment effect after the deadline. First, some subset of bureaucrats was induced by treatment to comply on time rather than late, as opposed to never complying at all. Second, PRONIED engages in costly follow-ups with non-compliers after the deadline by using a centralized call center and other means. Thus the slowly narrowing gap between treatment and control groups is, at least in part, a reflection of expensive manual efforts that the SMS campaign mitigates. Figure A1 in the Appendix presents the results for each treatment arm.

Figure 5 explores the dynamic of effects for the case of withdrawal of funds. We restrict the analysis to the case of withdrawal of 99% of maintenance funds. Except for a positive effect in May, months before the start of the program, all effects before the start of the SMS campaign are not distinguishable from zero. We do see evidence of positive impacts weeks after the beginning of the campaign. Due to restrictions on obtaining data from the National Bank after the end of the intervention, we are not able to evaluate whether the effect of the SMS campaign persists after the end of the intervention. However, results are in line with the estimates from the previous section. Figure A2 in the Appendix presents the results for each treatment arm.

[34] Figure S5 presents the results for the oversight report and Figure S6 the results for the approved expense report. Results for the different levels of the withdrawal of funds are reported in Figures S7, S8, S9 and S10.

6 Learning more about the effectiveness of non-monetary incentives

The results of our Benchmark Experiment provide support for the effectiveness of non-monetary incentives motivated by behavioral insights. Although all treatment arms are shown to improve compliance among civil servants, the social norm and the monitoring treatments in particular seem to play a strong role in improving outcomes while avoiding the potentially negative long-term effects of shaming and threatening civil servants with audits. With this in mind, we implemented a new large-scale field experiment in 2016 with the goal of further exploring the role of non-monetary incentives. Taking as a starting point the results of the Benchmark Experiment in 2015, we designed a new intervention to address the following questions:

(i) What types of social norms are most relevant?

(ii) Is making salient the social benefits of investing in school infrastructure an alternative way to improve compliance among civil servants?

(iii) Are the effects of this intervention persistent over time?

(iv) Does the duration of the SMS campaign matter?

Question (i) relates to the fact that we only exploited one particular type of social norm. Specifically, we considered a descriptive qualitative social norm where the reference group consisted of other maintenance civil servants in the same UGEL. However, social psychologists have proposed a distinction between descriptive and injunctive social norms (Cialdini and Goldstein, 2004, Rogers et al., 2018) and evidence suggests that individuals may react differently to them (Cialdini, 2007). We extend our analysis of social norms by incorporating treatments targeted to address the distinction between descriptive and injunctive social norms. We further explore the role of social norms by breaking down the descriptive social norm treatment into quantitative and qualitative versions, as well as modifying the reference group. We proceed in the same way to break down the injunctive social norm into two reference groups: parents and principals.[35]

Question (ii) relates to a growing literature that makes salient the social benefit of a task to influence recruitment and performance of civil servants (Ashraf et al., 2014; Ashraf et al., 2015). More specifically, we are interested in evaluating whether making salient PRONIED's mission can work as an extra source of intrinsic motivation for agents who perform social tasks (Besley and Ghatak, 2014). We vary the dimension of social benefit to consider messages that emphasize the importance of high-quality infrastructure for students' health (well-being social benefit treatment),

[35] As discussed in Mackie et al. (2015), there are 16 definitions of social norm in the social science literature, but all of them share some commonalities in terms of three critical elements: 1. A norm sets social expectations regarding one's behavior; 2. It depends on a group of reference; and 3. It is maintained by social influence. We experiment with variations in 1 and 2 in this intervention cycle.

for the pride of the school community (pride social benefit treatment), and for contributing to the students' learning (learning social benefit treatment).

Question (iii) addresses the important issue of whether civil servants can be induced to comply with the maintenance policies beyond a one-shot SMS campaign. Given the soft nature of our incentives, we should expect a rational agent to update his or her beliefs to realize that no penalty is enforceable to punish lack of compliance. Therefore, after the surprise element of the benchmark experiment in 2015 is gone, we might not expect any effect of the campaign on civil servants' compliance. On the other hand, if bureaucrats are forgetful or if SMS also serve to reduce problems of limited attention, effects should persist. We test this by comparing performance of civil servants according to their treatment status in the Benchmark Experiment to evaluate whether those who were previously exposed to the intervention in 2015 respond to the 2016 intervention.

Finally, question (iv) relates to understanding a critical component of a SMS campaign design: the treatment duration. We experimentally vary the number of SMS delivered to civil servants. One group receives a low-duration SMS campaign of 4 SMS delivered in a given period. A second group receives a high-duration campaign of 7 SMS that begins earlier but ends at the same time, at the end of the program cycle. Conceptually, one may expect a longer campaign to have a larger effect if it minimizes the risk of procrastination and inattention problems. On the other hand, if the number of messages is perceived to be too large, then it is possible that civil servants stop paying attention or treat them as "spam." Therefore, it is critical to take duration into account for an optimal campaign design.[36]

We designed a factorial block design to address these issues. The factorial components include the behavioral treatments (9 in total) and the duration (2 in total). The experiment was stratified on two dimensions: treatment status in the benchmark experiment (including a stratum for maintenance civil servants who are new to the sample) and the region.

The Online Appendix provides the details of the interventions, sample size, descriptive statistics and pre-treatment balance for the Follow-Up Experiment. Figure S1 presents the treatment arms, SMS content, and sample size for all treatment variants. Table S3 provides the SMS contents for all the messages delivered during the SMS campaign. Table S5 provides descriptive statistics for the follow-up experiment and Table S7 evaluates randomization balance.

6.1 Basic results for the Follow-up Experiment

We start by providing treatment effect estimates for the new campaign. We focus the analysis on the same compliance outcomes and the withdrawal of maintenance funds. We start by analyzing the overall impact of descriptive and injunctive social norms, as well as the social benefit without considering the variants in terms of reference group or whether the norm is quantitative or qualitative. We provide detailed results for these treatment variants below.

[36]The existence of saturation effects of informational campaigns have been documented in contexts such as voter mobilization (Buntaine et al., 2018) and parents' involvement in children's education (Cunha et al., 2017), but we know very little about the optimal design of SMS campaigns.

Basic results are reported in Table 5. All treatments have a positive and statistically significant impact on submission of the expense report (Column 1). The average SMS effect estimate is 1.73 percentage points (S.E. 0.536). Coefficient estimates for social norm treatments and the social benefit SMS are quantitatively similar. These represent a reduction of about 8% in the compliance gap. The effects of receiving a SMS are also positive and statistically significant for submission of the oversight report and the approval of the expense report, although effects are smaller and weaker in terms of statistical significance. The descriptive norm and the social benefit remain significant for the case of the oversight report (Column 2) and only the injunctive norm is significant for the approval of the expense report (Column 3). Results are largely unaffected after controlling for pre-treatment characteristics (Columns 4 through 6).

Several factors could explain the smaller treatment effects in this cycle, some of which we can consider as time-varying aggregate shocks as discussed in Rosenzweig and Udry (2016). Compliance in the control group was 6.5 percentage points higher than in the previous cycle, leaving less room to improve compliance. The maintenance cycle was also shorter, with an earlier deadline (August 31 instead of September 30), although even the short campaign covered approximately the same span of time as the Benchmark Experiment. The population of bureaucrats is substantially larger in the follow-up experiment, in part because MINEDU provided an additional database of cellphone numbers.

Table 6 presents the results for funds withdrawal from the National Bank accounts. We do not find effects of the SMS campaign on the withdrawal of any positive amount (Column 1), or at least 50%, 80%, or 95% of funds (Columns 2 through 4). The 0.9 percentage point effect on withdrawal of 99% of funds (Column 5) is significant at the 10% level, representing a reduction in the compliance gap of 8%. Results are robust to controlling for pre-treatment characteristics (Columns 6 to 10).

Figures 6 and 7 replicate the dynamic analysis of effects for the submission of the expense report and the withdrawal of at least 99% of maintenance funds, following the same strategy we implemented for the case of the Benchmark Experiment. In both cases, we find that no difference between treated and control civil servants exists before the beginning of the SMS campaign. Results become significant after the campaign starts. Figures A3 and A4 in the Appendix present the results for each treatment arm separately.

6.2 Additional results of the follow-up experiment

In this section, we explore the impact of the SMS campaign considering all treatment variations. In the case of the descriptive norms, we vary the norm in terms of whether quantitative information (reporting reference group compliance rate) or qualitative information (reporting that reference group compliance is high) is emphasized as well as varying the reference group. In particular, we decompose the descriptive social norm treatment into four categories:

(i) Qualitative social norm using the UGEL as the reference group

(ii) Quantitative social norm using the UGEL as the reference group

(iii) Qualitative social norm using the whole country as the reference group

(iv) Quantitative social norm using the whole country as the reference group

Regarding the injunctive social norm, we introduce variation in whether parents or principals are the reference group which defines whether a behavior (in this case, compliance with maintenance policies) is considered socially acceptable. Finally, we consider the social benefit treatments related to well-being, pride, and learning as described above.

Table A.1 presents results. Column 1 focuses on compliance with submission of the expense report. We find that both qualitative descriptive norms have a positive impact on submission of the expense report, with larger point estimates associated with the one that uses the UGEL as the reference group (point estimate of 2.57 and S.E of 0.85). Only the quantitative norm based on the whole country is significant at the 5% significance level. Regarding the injunctive social norms, only the one related to the principals is statistically significant (point estimate of 2.24 percentage points). For the social benefit treatment, the messages emphasizing pride and students' learning play a role in inducing compliance with larger point estimates than the average SMS.

The descriptive social norms have a very limited estimated impact on submission of the oversight report (Column 2) and approval of the expense report (Column 3). We find no statistically significant impacts, except for the case of the quantitative norm based on UGEL as a reference group for the case of compliance with the oversight report and the quantitative norm based on the whole country as reference group for approval of the expense report. The injunctive norm based on principals also induces compliance in the submission of both reports as well as the social benefit treatments based on pride and students' learning. These results are robust to the inclusion of pre-treatment controls (Columns 4 to 6).

The results for the case of withdrawal of maintenance funds are reported in Table A2. The point estimate for the qualitative descriptive norm suggests that using the whole country as the reference group may be very effective in inducing withdrawal of maintenance funds. This is true for all levels of withdrawal higher than 50%. The quantitative descriptive norms also induce withdrawal of maintenance funds for the case of 95% and 99% levels. Both injunctive norms induce withdrawal of maintenance funds for more than 80% levels. On the other hand, the social benefit message related to students' learning is also effective in inducing withdrawal of maintenance funds for levels higher than 50%. These results are also robust to the inclusion of pre-treatment controls (Columns 6 to 10).

6.3 Effect persistence

Here we evaluate whether being exposed to a treatment arm in the Benchmark Experiment affects civil servants' response in the follow-up experiment. Due to space constraints, we focus on

compliance with the submission of the expense report. We compare the effects of receiving a SMS conditional on treatment status in the Benchmark Experiment. Results are presented in Table 7.

We compare three scenarios. In the first (Panel A), we estimate the impact of the SMS campaign for those who were part of the control group in 2015 but become part of a treatment group in 2016 with respect to a pure control group (those not exposed in 2015 or 2016). We find a positive impact of about 2.3 percentage points (significant at the 10% significance level) for the case of submission of the expense report (Column 1). This can be interpreted as evidence that knowledge of the 2015 intervention by members of the control group in 2015 did not prevent them from responding to the 2016 intervention. In the second scenario (Panel B), we compare the response (again vs. pure control) of those who were exposed to the intervention in both 2015 and 2016, finding a positive impact of the SMS campaign in 2016. This suggests that previous exposure to the SMS campaign does not affect the ability of the program to induce compliance among civil servants. In the final scenario (Panel C), we consider the case of those exposed to treatment in 2015 but assigned to the control group in 2016. If treatment effects persist over time, we should expect for this group to have a higher compliance rate than those in the pure control group. We find no evidence in favor of this hypothesis. Estimates are not only statistically insignificant, but also negative and close to zero. We thus have no evidence of the effect of past treatment on current behavior, suggesting that civil servants do not develop specific knowledge about the intervention that affects their future behavior.

Columns 2 and 3 present the results for submission of the oversight and approved expense report. Results are robust to controlling for pre-treatment characteristics. Results of this section suggest that the effects of a SMS campaign do not persist over time and that this tool can be used to influence compliance on a regular basis. Civil servants either do not update their beliefs regarding the credibility of the text messages' content, or messages help them to overcome another behavioral barrier such as limited attention.

6.4 Campaign duration

This section explores the role of campaign duration. Table A.3 shows the effect of the number of messages delivered during the behavioral intervention on WASICHAY outcomes. The long duration treatment group received 7 messages and the short duration treatment only 4. Column 1 shows that both SMS campaigns induce compliance with submission of the expense report, although the short duration SMS campaign has a higher point estimate (1.93 versus 1.53 percentage points, not statistically different at conventional levels). The short duration SMS campaign also had statistically significant effects on compliance with delivery of the oversight and approval of the expense reports (Columns 2 and 3), an effect that is insignificant for the long campaign although the effects are not statistically distinguishable from each other. These results are robust to controlling for pre-treatment characteristics (Columns 4 to 6).

Table A.4 shows the effects of the two campaigns on withdrawal of maintenance funds. We

find that the short campaign induced a significant increase in the withdrawal of at least 95% and 99% of the maintenance funds. This is not the case for the high-intensity campaign, although again, estimated impacts are not statistically distinguishable from each other at conventional levels. Results are similar when pre-treatment controls are included. Taken together, the evidence suggests that the short campaign was likely at least as effective as the long one. One implication is that, when designing a SMS-based intervention, policymakers should consider the possibility that a more limited campaign could save financial and human resources without sacrificing impact on behaviors of interest.

7 External validity

In previous sections, we have shown that non-monetary incentives based on behavioral insights work as a tool to motivate civil servants to comply with national policies. These effects were found with large-scale experiments that cover almost all of the population of maintenance civil servants for which a cellphone was registered with MINEDU. Therefore, our design does not suffer from the standard external validity case in which an innovation is tested in a limited number of geographical or political units in a given country. The results described in the previous section are externally valid for the maintenance civil servants in the country.

However, the results of our experiment are not necessarily externally valid for all civil servants. This is not the typical external validity concern that has concerned economists in the recent scholarship that uses field experiments in developing countries. Although it is almost impossible to run an experiment to address external validity for all civil servants, we take one step forward by running an additional field experiment in a different population of civil servants to shed light on the applicability of our intervention in other settings. Although we recognize that running additional experiments does not exhaust all relevant dimensions of external validity, we believe this exercise illuminates relevant issues to consider when it comes to understanding the applicability of our results to other settings. In addition, by implementing the Follow-Up Experiment in 2016 with the population of civil servants, we have also addressed the external validity of our results in the presence of aggregate time-specific shocks (Rosenzweig and Udry, 2016).

In this section, we study whether the effects of a version of the SMS campaign are similar to the ones found with maintenance civil servants. We partnered with the National Program CUNA MAS, an early childhood development program, to implement a SMS campaign to motivate compliance among bureaucrats in charge of a family support service. This campaign was designed to incorporate the lessons learned from the 2015 and 2016 experiments with MINEDU. Given the effects found with the social norm and monitoring treatments, we designed a SMS campaign based on these two behavioral contents.[37]

The outcome of interest in this external validity experiment is compliance with reporting of

[37] We consider a qualitative descriptive social norm using the Territorial Unit as reference group. This resembles the social norm treatment used for the benchmark experiment in 2015. See the Online Appendix for more details.

service delivery. CUNA MAS requires updated information about the delivery of services (home visits) as well as program beneficiaries' progress on a monthly basis. This information is used to plan service delivery for the next period as well as to update the beneficiary list to incorporate new families. Lack of compliance with the submission of this information makes it difficult for this program to respond to its beneficiaries' needs. This opens the room for an intervention that can induce compliance in a cost-effective manner.

The Online Appendix presents the technical details of the implementation of this external validity experiment and a set of analyses to evaluate the internal validity of our research design.[38] Due to space restrictions, we focus on the basic results of the intervention, presented in Table 8. We aggregate the results for the five months (from September 2016 to January 2017) in which the intervention was in place. As in the case of PRONIED, all civil servants with access to a tablet assigned by the program were exposed to the intervention (1,093 across the country). The necessary information to evaluate the impact of the intervention was obtained from CUNA MAS administrative records. The outcome of interest is the percentage of scheduled field visits that were actually reported by the supervisor by the monthly deadline.

Overall, the results suggest that the monitoring treatment is the most effective tool to induce compliance in this population. For instance, in Column 1, the estimated monitoring SMS effect is 4.8 percentage points and is significant at the 10% level. Considering a compliance level of 70% for the control group, the previous estimate represents a 16% reduction in terms of the compliance gap. Point estimates remain unchanged after including control variables in the basic specification (Column 2) but standard errors are lower and the effect is significant at the 5% level. Further controlling for baseline outcomes increases the point estimates (5.5 percentage points, significant at the 1% level), representing a reduction in terms of the compliance gap of 18% (Column 3). We also analyze an additional specification in which we drop the October and December months from the sample due to implementation issues. October was dropped because tablets' operating systems were updated and civil servants' reporting duties were not enforced by program administrators. December was excluded because SMS were not sent due to the holiday season. Estimated coefficients slightly increase (Column 4), suggesting that these results are robust to the exclusion of periods related to implementation issues.

We interpret these results as evidence that civil servants' characteristics matter for understanding the applicability of the lessons of this set of interventions. In particular, we propose the hypothesis that tenure differences might explain the lack of evidence in favor of social norms. Whereas public officials at MINEDU typically have long-term contracts and are highly unionized, CUNA MAS bureaucrats are hired using a variety of short-term contracts. Since the activities they need to perform do not require of a specific set of skills, they tend to have lower qualifications than

[38] Figure S3 describes the timing of the intervention and Figure S4 provides the details of the treatments, sample size and examples of the text messages delivered in the intervention. Table S4 presents the contents of the text messages delivered during the SMS campaign. Table S6 provides the descriptive statistics and Table S8 evaluates the randomization balance.

officials at MINEDU and higher levels of turnover. Consequently, they are less sensitive to the perceptions that their colleagues may have regarding their work. This may explain why appealing to social norms mean very little to them. On the other hand, monitoring works better because it is perceived as a tool that can affect their chances of keeping their jobs.

8 Additional results

In this section, we present additional results regarding heterogeneous impacts and some suggestive evidence on the impact of the intervention on infrastructure quality.

8.1 Heterogeneous effects

This section explores whether the intervention caused differential impacts with respect to different covariates of interest. We had no prior beliefs about particular dimensions of heterogeneity and did not stratify the design accordingly. This analysis should be considered exploratory and descriptive. The analysis is restricted to exploring the heterogeneous impacts of the SMS campaign on expense report submission (Table A5) and the withdrawal of 99% of the maintenance funds (Table A6) in the Benchmark Experiment. We focus on heterogeneities with respect to school size, the assigned budget, assigned budget per capita, and an indicator of rurality.

We find no significant evidence of heterogeneous effects, although suggestive evidence is found in some dimensions. For instance, we find that the interaction between the assigned budget per student and the treatment is positive, although the coefficient is close to 0 (Column 6, Table A5). This result is robust to controlling for pre-treatment characteristics (Column 14, Table A5). We find no evidence for the case of school size and assigned budget.

We also find some weak evidence of heterogeneous effect for school size in explaining the withdrawal of up to 99% of maintenance funds (Column 2, Table A6), but this result is not robust after controlling for pre-treatment characteristics (Column 10, Table A6). We do find some evidence of heterogeneous impact for the case of the assigned budget per student, but the coefficient is quite small (0.006). We do not find evidence of heterogeneous impacts for the remaining dimensions considered in the analysis.

8.2 Effects on maintenance expenditures

The results of the Benchmark and Follow-Up experiments suggest that the intervention is inducing compliance with the withdrawal of funds and delivery of the expense report. In this section, we analyze whether the intervention is inducing civil servants to use maintenance funds to invest in the infrastructure categories prioritized in the technical form they have filed. Recall that, before the intervention, the maintenance civil servant–along other members of the maintenance committee–define the investment priorities and prepare a technical form with a budget to be approved by the

UGEL. We want to test whether the intervention affected the way funds are spent across investment categories.

This is an important issue because of the concern that the intervention, by motivating civil servants to withdraw and spend funds, may also have induced an increase in malfeasance and corruption. We explore this matter by comparing the planned expenditures (as registered in the technical form) against the executed expenditures (as reported in the expense report) in the Follow-up Experiment, for which we have the appropriate administrative data. Significant departures from the planned expenditures can be interpreted as a signal that the SMS campaign affected the way civil servants used the maintenance funds in ways that were inconsistent with the wishes of the school community. Although this is not necessary or sufficient evidence of malfeasance or corruption, it could be certainly consistent with them.

Table 9 reports the results of an empirical exercise where the dependent variable is the difference between the executed and planned investment categories, measured in PEN. We consider the most common investment categories, including repair of ceilings, floors, sanitary facilities, walls, doors, and windows. We also consider electrical installations along with repair and replacement of furniture and other school supplies. The first thing to note is that the average spending gap in the control group is PEN -77 (Column 1, bottom panel). Although this amount is small (about $20 USD), this means that they spent less than planned. On the other hand, those in the treatment group spent PEN 33 more than the control group, implying that the intervention is inducing higher expenditures but not enough to arrive at a positive net effect. In sum, treated civil servants do spend more but still below their planned expenditure. Therefore, it does not seem that the intervention is causing a large deviation from planned expenditures.

Results are similar when we look into investment categories. With the exception of school supplies, we find no differences in most investment categories. Coefficients are not only statistically insignificant, but also small in magnitude. In the case of school supplies, there is a positive effect (PEN 5.1, significant at the 10% significance level), but small and still reflecting a negative net effect (control mean of PEN -13). Taking into account the fact that the expense report has to be backed up with invoices, receipts and similar evidence, the possibility that the intervention is inducing patterns of investment consistent with corruption seems unlikely.

8.3 Effects on infrastructure quality

In this section, we discuss whether the intervention had an impact on infrastructure quality. Given the nature of the intervention, one may expect that the intervention caused an increase in the quantity and quality of the infrastructure and furniture. An alternative view would suggest that the SMS campaign could have negatively affected the quality of the infrastructure and furniture by inducing substitution toward spending on items that are faster to implement.

To address these issues, we exploit information about principals' and teachers' perceptions regarding infrastructure and furniture quality, using the infrastructure module included in the

SEMAFORO survey. In this module, teachers and principals are required to evaluate whether they consider the quality of the infrastructure (including walls, ceilings, floors, etc.) as well as the quality of the classroom furniture (including desk, chairs, boards, etc.) to be either good, regular, or bad.[39] We use data for the 2016 version of the survey.[40]

In Table 10, we explore the effect of the SMS campaign on infrastructure quality as reported by the teachers. We created a dummy variable equal to 1 if the surveyed teacher considered the quality of a specific item of infrastructure to be bad. Responses were normalized to the 0-100 scale to facilitate interpretation. The items of infrastructure under analysis are walls, ceilings, floors, windows, doors and bathroom sinks.

We find no evidence that the SMS campaign affected the quality of most of the infrastructure items, except for the case of bathroom sinks (Column 5). In this case, we find suggestive evidence that the intervention caused a reduction of those teachers reporting that the quality of the bathroom sink is bad (coefficient of -2.9, significant at the 10% level). This result is robust to controlling for pre-treatment characteristics. These results are driven by the social benefit treatment.

The lack of evidence of impacts on quality is not surprising, given the nature of the intervention. Maintenance funds are relatively small while significant changes in infrastructure quality are expensive. In addition, the proportion of teachers who report that a particular infrastructure item is of bad quality is typically less than 5%, with the exceptions of floors (8%) and doors (6%). The fact that an impact is found precisely in the case of bathroom sinks, which are considered of bad quality for most teachers (59%), could be consistent with a scenario where maintenance civil servants are targeting funds to solve an issue that is considered relevant for teachers in a context where budget constraints are important. In this way, we interpret this result as suggestive evidence that the SMS campaign is not only inducing maintenance civil servants to comply with the policy, but also that they are doing so in a way that improves the quality of those items that are the most relevant for teachers, when the social benefit of the use of maintenance funds is emphasized.

We also explore the effect of the SMS campaign on the quality of furniture (Table A7). The fraction of teachers that report that the quality of furniture is bad is quite low for all furniture items under consideration. These include teacher and student desks and chairs, boards and cupboards. We find no evidence of impacts of the SMS campaign for any of these items. Finally, we analyze the role of the SMS campaign in affecting the stock of infrastructure (Table A8). The intervention has no apparent impact on whether the school has a toilet or the number of them, but it does affect the number of sinks (Column 3). Results are robust to controlling for pre-treatment characteristics. This result is consistent with those of Table 10, showing that the intervention not only affected the quality but also the stock of bathroom sinks.

[39] All teachers in all grades and sections were required to provide an answer to this question. This implies that for large schools several responses are available because they typically offer multiple grades in primary and secondary levels. Given that most schools in the country offer primary and offer only one section per grade, we restrict the analysis to teachers in charge of the first section in first grade to ensure comparability.

[40] The SEMAFORO survey was also carried out in 2015, but the survey instruments were heavily revised in 2016. To avoid comparability issues, we restrict the analysis to 2016.

8.4 Cost-effectiveness analysis

In this section, we report a simple cost-effectiveness exercise for the Benchmark Experiment. The goal is to estimate the additional amount of funds accounted for in the expense report per dollar spent on the intervention, measured at the expense report deadline. Because MINEDU already has infrastructure for sending SMS, the primary monetary cost incurred is the per-message cost to transmit the SMS over the cellular network. While we are not privy to the terms of MINEDU's contract with its service provider, the price per SMS on the bulk market is approximately PEN 0.07 ($0.022). A total of 57,860 SMS were sent in the experiment, leading to an estimated cost of PEN 4,050 ($1,273). In order to estimate the additional funds reported as a result of the intervention, we first estimate equation 1 with amount declared in the expense report as the dependent variable and a single dummy variable for exposure to any SMS treatment arm. The estimated coefficient is PEN 211.51 ($66.10). A total of 17,545 civil servants were treated, so the estimated additional amount declared is 211.51 $\times$ 17,545 = PEN 3,710,942.95 ($1,159,670). Dividing this amount by the total cost, we arrive at PEN 916 reported per PEN 1 spent on SMS (i.e. $916 per $1 spent).[41]

To account for labor costs associated with the programming and sending of the SMS, we assume that two hours are spent for each of the five waves of SMS messages. Assuming a labor cost of PEN 60/hour, this adds PEN 600 ($188) to the total campaign cost. Note that this cost is independent of the number of recipients, so scaling to the full population will decrease the effect of labor costs on this cost-effectiveness measure. For this intervention, the estimated cost-effectiveness incorporating labor costs is PEN 798 per PEN 1 spent (i.e. $798 per $1 spent). While this cost-effectiveness ratio appears very large, the efficacy of the intervention depends on how much PRONIED values the timely reporting of these funds. This is difficult to assess, but one indicator that PRONIED found the intervention effective is that they have subsequently scaled it up to the full population in every maintenance cycle. An additional benefit to PRONIED is that, by inducing timely reporting of expenditures, their call center can reduce the number of calls that it makes to delinquent civil servants, potentially resulting in significant additional cost savings.

9 Conclusion

In this paper, we conducted a set of field experiments to provide evidence on the potential effectiveness of non-monetary incentives, based on behavioral insights, in inducing compliance among civil servants in a setting where the government lacks the capacity to monitor and punish them. We find that these incentives are a very cost-effective strategy to induce compliance. Our results highlight the importance of carefully designed non-monetary incentives as a tool to improve civil servant performance when the state lacks institutional mechanisms to enforce compliance. A wide variety of message contents were found to be effective, reducing the compliance gap significantly at low cost. Furthermore, the intervention was found to be effective even when civil servants have

[41] The standard error of this estimate is PEN 294, which accounts for estimation error from the regression.

already been exposed to it in the previous year, indicating that sending SMS reminders as a matter of policy may have benefits over a course of many years. A very short campaign consisting of only four messages was apparently as effective as a longer one, indicating the potential for implementing simple, concentrated interventions that improve civil servant compliance at low cost.

In our External Validity Experiment, we explore the effects of a version of the Benchmark Experiment in a setting in which civil servants have different characteristics, in particular differences in tenure. Working with CUNA MAS, an early childhood development program, we find that the monitoring treatment is relevant whereas social norms no longer play a role. We interpret this result as evidence of the role of the type of labor contract under which the civil servant is governed, since maintenance civil servants are typically hired under long-term contracts but CUNA MAS employees have temporary contracts. These results show that the institutional setting of the intervention can interact with the type of non-monetary incentives being implemented, such that policy designers need to be familiar with the civil servants they are targeting and the environment in which they operate.

Further research is needed to evaluate the role of other types of behavioral insights to increase civil servant compliance. Most of our treatments have been developed around the notions of limited attention and social norms, but they are far of covering the large range of available options to design SMS campaigns based on behavioral insights. There is also room to explore alternative means to deliver the interventions beyond text messages. Furthermore, there is substantial room to improve targeting of interventions toward civil servants who are predicted to be non-compliant or for whom treatment is predicted to be most effective. Such approaches, assisted by machine learning techniques made possible given the scale of large interventions, may both save monetary resources and avoid taxing the time and attention of civil servants who were going to comply anyway. Despite this, we believe that the most important result of this paper is to show that using behavioral insights via SMS campaigns can be a powerful, cost-effective, scalable tool to induce compliance among civil servants.

References

Aker, J., P. Collier, and P. Vicente (2017). Is information power? Using mobile phones and free newspapers during an election in Mozambique. *Review of Economics and Statistics 99*(2), 185–200.

Al-Ubaydli, O., J. List, and D. Suskind (2017). What can we learn from experiments? Understanding the threats to the scalability of experimental results. *American Economic Review: Papers and Proceedings 107*(5), 282–286.

Alcott, H. and T. Rogers (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review 104*(10), 3003–3037.

Allcott, H. (2015). Site selection bias in program evaluation. *Quarterly Journal of Economics 130*(3), 1117–1165.

Andreoni, J., M. Callen, Y. Khan, K. Jaffar, and C. Sprenger (2016). Using preference estimates to customize incentives: An application to polio vaccination drives in pakistan. NBER Working Paper No. 22019.

Andrews, I. and E. Oster (2017). Weighting for external validity. NBER Working Paper 23826.

Angrist, J., G. Imbens, and D. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association 91*, 444–455.

Ariely, D. and K. Wertenbroch (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science 13*(3), 219–224.

Ashraf, N., O. Bandeira, and K. Jack (2014). No margin, no mission? A field experiment on incentives for public service delivery. *Journal of Public Economics 120*, 1–17.

Ashraf, N., O. Bandiera, and S. S. Lee (2015). Do-gooders and go-getters: Career incentives, selection, and performance in public service delivery. HBS and NBER, Department of Economics and STICERD, Harvard Medical School and Harvard Business School.

Azulai, M., O. Bandiera, F. Blum, H. Kleven, E. La Ferrara, G. Padr, and C. M. Tejada (2014). State effectiveness, growth, and development. *International Growth Centre*.

Banerjee, A., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukerji, M. Shotland, and M. Walton (2017). From proof of concept to scalable solutions, with an application. *Journal of Economic Perspectives 31*(4), 73–102.

Banerjee, A., S. Chassang, and E. Snowberg (2017). Decision theoretic approaches to experiment design and external validity. In E. Duflo and A. Banerjee (Eds.), *Handbook of Field Experiments*, Volume I, pp. 142–174. Elsevier B.V.

Banerjee, A., R. Chattopadhyay, E. Duflo, D. Keniston, and N. Singh (2014). Improving police performance in Rajasthan, India: Experimental evidence on incentives, managerial autonomy and training. NBER Working Paper No. 17912.

Banuri, S., S. Dercon, and V. Gauri (2017). Biased policy professionals. World Bank Policy Research Working Paper 8113.

Banuri, S. and P. Keefer (2016). Pro-social motivation, effort and the call to public service. *European Economic Review 83*, 139–164.

Behavioral Insights Team (2017). An exploration of potential behavioural biases in project delivery in the Department for Transport. Technical report, Behavioral Insights Team.

Bertrand, M., R. Burgess, A. Chawla, and G. Xu (2017). The glittering prizes: Career incentives and bureaucrat performance. University of Chicago, London School of Economics, International Growth Centre, Indian Administrative Service, University of California at Berkeley.

Besley, T. and M. Ghatak (2014). Solving agency problems: Intrinsic motivation, incentives, and productivity. Background paper for the World Development Report 2015.

Besley, T. and T. Persson (2009). The origins of state capacity: property rights, taxation, and politics. *American Economic Review 99*(4), 1218–1244.

Besley, T. and T. Persson (2010). State capacity, conflict, and development. *Econometrica 78*, 1–34.

Besley, T. and T. Persson (2011). *The Pillars of Prosperity. The Political Economics of Development Clusters.* Princeton University Press.

Best, M. C., J. Hjort, and D. Szakonyi (2017). Individuals and organizations as sources of state effectiveness, and consequences for policy. NBER Working Paper No. 23350.

Bhargava, S. and D. Manoli (2015). Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment. *American Economic Review 105*(11), 3489–3529.

Bobonis, G., L. Camara, and R. Schwabe (2016). Monitoring corruptible politicians. *American Economic Review 106*(8), 2371–2405.

Bruhn, M. and D. McKenzie (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics 1*(4), 200–232.

Buntaine, M., R. Jablonski, D. Nielson, and P. Pickering (2018). SMS text on corruption help Ugandan voters hold elected councillors accountable at the polls. *Proceedings of the National Academy of Sciences 115*(26), 6668–6673.

Burgess, S., C. Propper, M. Ratto, and E. Tominey (2017). Incentives in the public sector: Evidence from a government agency. *The Economic Journal 127*, 117–141.

Burgess, S. and M. Ratto (2003). The role of incentives in the public sector: Issues and evidence. *Oxford Review of Economic Policy 19*(2).

Bursztyn, L., A. Gonzalez, and D. Yanagizawa-Drott (2018, June). Misperceived social norms: Female labor force participation in Saudi Arabia. NBER Working Paper 24736.

Callen, M., S. Gulzar, A. Hasanain, M. Khan, and A. Rezaee (2018). Data and policy decisions: Experimental evidence from Pakistan. Working Paper, UC San Diego.

Callen, M., S. Gulzar, A. Hasanain, and Y. Khan (2014). The political economy of public employee absence: Experimental evidence from Pakistan. University of California, New York University, Lahore University of Management Sciences, International Growth Centre.

Calzolari, G. and M. Nardotto (2017). Effective reminders. *Management Science 63*(9), 2915–2932.

Celhay, P., P. Gertler, P. Giovagnoli, and C. Vermeersch (2015). Long run effects of temporary incentives on medical productivity. NBER Working Paper 21361.

Chaudhury, N., J. Hammer, M. Kremer, K. Muralidhara, and F. H. Rogers (2006). Missing in action: Teacher and health worker absence in developing countries. *Journal of Economic Perspectives 20*(1), 91–116.

Chong, A., D. Karlan, J. Shapiro, and J. Zinman (2013). (Ineffective) messages to encourage recycling: Evidence from a randomized evaluation in Peru. *World Bank Economic Review 29*(1), 180–206.

Cialdini, R. B. (2007). Descriptive social norms as underappreciated sources of social control. *Psychometrika 72*(2), 263.

Cialdini, R. B. and N. J. Goldstein (2004). Social influence: Compliance and conformity. *Annual Review of Psychology 55*, 591–621.

Cingolani, L., K. Thomsonn, and D. De Crombrugghe (2015). Minding Weber more than ever? The impacts of state capacity and bureaucratic autonomy on development goals. *World Development 72*, 191–207.

Cunha, N., G. Lichand, R. Madeira, and E. Bettinger (2017). What is it about communicating with parents? Technical report, Stanford University.

Dal Bo, E. and F. Finan (2016). At the intersection: A review of institutions in economic development. University of California, Berkeley.

Dal Bo, E., F. Finan, N. Li, and L. Schechter (2018). Government decentralization under changing state capacity: Experimental evidence from Paraguay. NBER Working Paper 24879.

Dal Bo, E., F. Finan, and M. A. Rossi (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. *Quarterly Journal of Economics*, 1169–1218.

Davis, J., J. Guryan, K. Hallberg, and J. Ludwig (2017). The economics of scale-up. NBER Working Paper 23925.

De La O, A. and F. Martel (2015). Can intrastate accountability reduce local capture? Results from a field experiment Mexico. Working Paper, Yale University.

Delmas, M. and N. Lessem (2014). Saving power to conserve your reputation? The effectiviness of private versus public information. *Journal of Environmental Economics and Management 67*(3), 353–370.

Deserranno, E. (2016). Financial incentives as signals: Experimental evidence from the recruitment of health promoters. Northwestern University, Kellogg School of Management.

Dhaliwal, I. and R. Hanna (2017). The devil is in the details: The successes and limitations of bureaucratic reform in India. *Journal of Development Economics 124*, 1–21.

Dixit, A. (2002). Incentives and organizations in the public sector: An interpretative review. *Journal of Human Resources 37*(4), 696–727.

Duflo, E., M. Greenstone, R. Pande, and N. Ryan (2013). Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India. *Quarterly Journal of Economics 128*(4), 1499–1545.

Duflo, E., R. Hanna, and S. P. Ryan (2012). Incentives work: Getting teachers to come to school. *American Economic Review 102*(4), 1241–1278.

Englich, B., T. Mussweiler, and F. Strack (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on expert's judicial decision making. *Society for Personality and Social Psychology 32*(2), 188–200.

Eyal, N. (2014). Nudging by shaming, shaming by nudging. *International Journal of Health Policy and Management 3*(2), 53.

Ferraro, P. and J. Miranda (2013). Heterogeneous treatment effects and mechanisms in information-based environmental policies: Evidence from a large-scale field experiment. *Resource and Energy Economics 35*, 356–379.

Ferraz, C. and F. Finan (2008). Exposing corrupt politicians: The effects of Brazil's publicly released audits on electoral outcomes. *Quarterly Journal of Economics 123*(2), 703745.

Finan, F., B. Olken, and R. Pande (2017). The personnel economics of the developing state. In A. Banerjee and E. Duflo (Eds.), *Handbook of Field Experiments*, Volume 2, Chapter 6, pp. 467–514. Elsevier B.V.

Foxcroft, D., M. Moreira, N. Almedida, and L. Smith (2015). Social norms information for alcohol misuse in university and college students. Technical Report 2, Cochrane Database of Systematic Reviews.

Frey, B. and S. Meier (2004). Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment. *American Economic Review 94*(5), 1717–1722.

Gerber, A., D. Green, and C. Larimer (2010). An experiment testing the relative effectiveness of encouraging voter participation by inducing feelings of pride or shame. *Political Behavior 32*(3), 409–422.

Gerber, A. and T. Rogers (2009). Descriptive social norms and motivation to vote: Everybody's voting and so should you. *The Journal of Politics 71*(1), 178–191.

Gertler, P., M. Shah, M. L. Alzua, L. Cameron, S. Martinez, and S. Patil (2015). How does health promotion work? Evidence from the dirty business of eliminating open defecation. NBER Working Paper 20997.

Glewwe, P., N. Ilias, and M. Kremer (2010). Teacher incentives. *American Economic Journal: Applied Economics 2*(3), 205–227.

Grindle, M. (2012). *Jobs for the Boys. Patronage and the State in Comparative Perspective*. Harvard University Press.

Hallsworth, M., M. Egan, J. Rutter, and M. Julian (2018). Behavioural government. Using behavioural science to improve how governments make decisions. The Behavioural Insights Team.

Hallsworth, M., J. A. List, R. D. Metcalfe, and I. Vlaev (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics 148*, 14–31.

Halpern, D. (2016). *Inside the Nudge Unit. How Small Changes Can Make A Big Difference*. WH Allen.

Hansen, W. and J. Graham (1991). Preventing alcohol, marijuana, and cigarette use among adolescents: Peer pressure resistance training versus establishing conservative norms. *Preventive Medicine 20*(3), 414–430.

Hasnain, Z., N. Manning, and J. H. Pierskalla (2014). The promise of performance pay? Reasons for caution in policy prescriptions in the core civil service. *World Bank Research Observer 29*, 235–264.

Karachiwalla, N. and A. Park (2017). Promotion incentives in the public sector: Evidence from Chinese schools. *Journal of Public Economics 146*, 109–128.

Karlan, D. and M. McConnell (2014, October). Hey look at me: The effect of giving circles on giving. *Journal of Economic Behavior & Organization 106*, 402–412.

Karlan, D., M. McConnell, S. Mullainathan, and J. Zinman (2016). Getting to the top of mind: How reminders increase saving. *Management Science 62*(12), 3393–3411.

Karlan, D., M. Morten, and J. Zinman (2012). A personal touch: Text messaging for loan repayment. NBER Working Paper 17952.

Khan, A. Q., A. I. Khwaja, and B. A. Olken (2016). Tax farming redux: Experimental evidence on performance pay for tax collectors. *Quarterly Journal of Economics*, 219–271.

Khan, A. Q., A. I. Khwaja, and B. A. Olken (2018). Making moves matter: Experimental evidence on incentivising bureaucrats through performance-based transfers. NBER Working Paper 24383.

Kowalski, A. (2018). How to examine externality validity within an experiment. NBER Working Paper 24834.

Langfeldt, L. (2004). Expert panels evaluating research: Decision-making and sources of bias. *Research evaluation 13*(1), 51–62.

Lee, D. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies 76*(3), 10711102.

Lewis-Faupel, S., Y. Neggers, B. Olken, and R. Pande (2016). Can electronic procurement improve infraestructure provision? Evidence from public works in India and Indonesia. *American Economic Journal: Economic Policy 8*(3), 258–283.

Lynch, J., R. Mowrey, G. Nesbitt, and D. O'Neill (2004). Risky business: Misperceived norms of sexual behavior among college students. *NASPA Journal 42*(1), 21–35.

Mackie, G., F. Moneti, H. Shakya, and E. Denny (2015). What are social norms? How are they measured? UNICEF/UC San Diego.

McCaffrey, M. (2014). Incentives and the economic point of view: The case of popular economics. *The Review of Social and Economic Issues 1*(1), 71–87.

Meyer-Sahling, J., C. Schuster, and K. Mikkelsen (2018). Civil service management in developing countries: what works? Evidence from a survey with 23,000 civil servants in Africa, Asia, Eastern Europe and Latin America. Technical report, UK Department for International Development.

Miller, D. and D. Prentice (2016). Changing norms to change behavior. *Annual Review of Psychology 67*, 339–361.

Molina, E., L. Carella, A. Pacheco, G. Cruces, and L. Gasparini (2017). Community monitoring interventions to curb corruption and increase access and quality in service delivery: A systematic review. *Journal of Development Effectiveness 9*, 462–499.

Muralidharan, K. and P. Niehaus (2017). Experimentation at scale. *Journal of Economic Perspectives 31*(4), 103–124.

Muralidharan, K., P. Niehaus, and S. Sukhtankar (2016). Building state capacity: Evidence from biometric smartcards in India. *American Economic Review 106*(10), 2895–2929.

Olken, B. A. (2007). Monitoring corruption: Evidence from a field experiment in Indonesia. *Journal of Political Economy 115*(2), 200–249.

Pepinsky, T., J. Pierskalla, and A. Sacks (2017). Bureaucracy and service delivery. *Annual Review of Political Science 20*, 249–268.

Peters, J., J. Langbein, and G. Roberts (2018). Generalization in the tropics-development policy, randomized controlled trials, and external validity. *The World Bank Research Observer 33*(1), 34–64.

Rogers, T., N. Goldstein, and C. Fox (2018). Social mobilization. *Annual Review of Psychology 69*, 357–381.

Rosenzweig, M. and C. Udry (2016). External validity in a stochastic world. NBER Working Paper 22449.

Savje, F., P. M. Aronow, and M. G. Hudgens (2017). Average treatment effects in the presence of unknown interference. *arXiv preprint arXiv:1711.06399*.

Savoia, A. and K. Sen (2014). Measurement, evolution, determinants, and consequences of state capacity: a review of recent research. *Journal of Economic Surveys 29*, 441–458.

Schultz, P. W. (1999). Changing behavior with normative feedback interventions: A field experiment on curbside recycling. *Basic and Applied Social Psychology 21*(1), 25–36.

Sonntag, A. and D. Zizzo (2015). On reminder effects, drop-outs and dominance: Evidence from an online experiment on charitable giving. *PLOS One*.

Stewart, D. and G. Stasser (1995). Expert role assignment and information sampling during collective recall and decision making. *Journal of Personality and Social Psychology 69*(4), 619–628.

Thaler, R. and C. Sunstein (2008). *Nudge. Improving Decisions About Health, Wealth, and Happiness*. Penguin Books.

Vervloet, M., A. Linn, J. van Weert, D. de Bakker, M. Bouvy, and L. van Dijk (2012). The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: A systematic review of the literature. *Journal of the American Medical Informatics Association 19*, 696–704.

Voigt, K. (2013). "If you smoke, you stink". Denormalisation strategies for the improvement of health-related behaviours: The case of Tobacco. In D. Strech, I. Hirschberg, and G. Marckmann (Eds.), *Ethics in Public Health and Health Policy. Concepts, Methods, Case Studies.*, pp. 47–62. Springer.

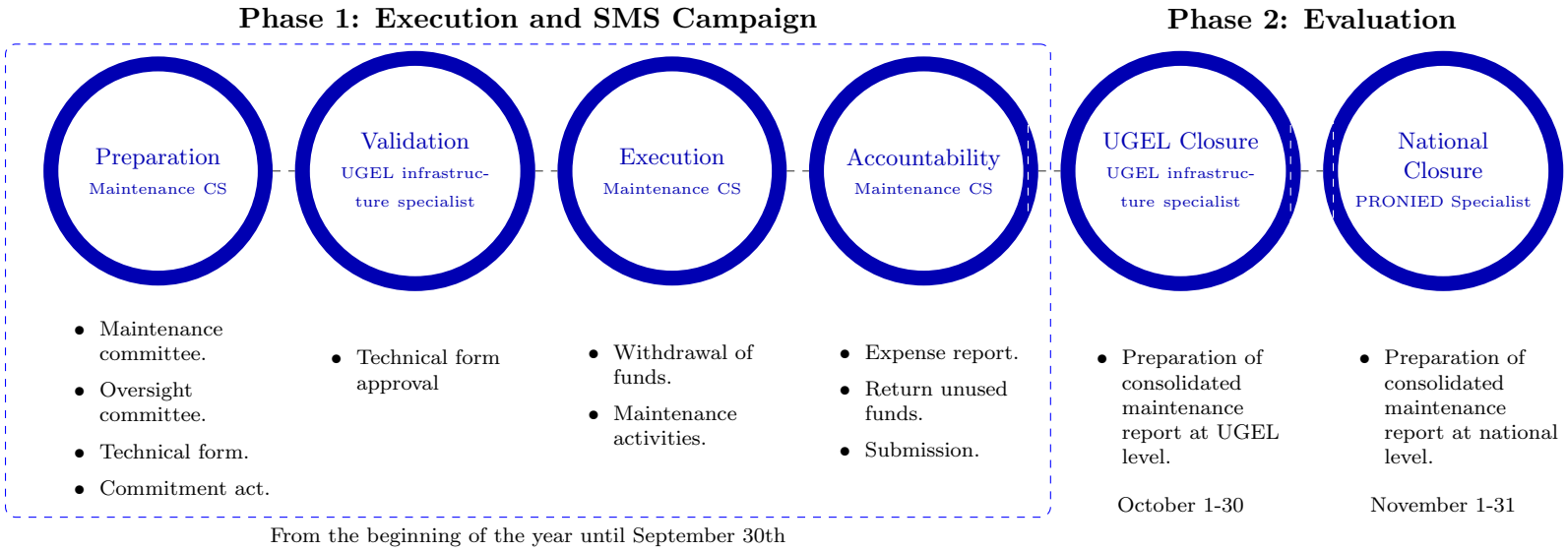
WHO (2009). Changing cultural and social norms that support violence. Technical report, World Health Organization.

World Bank (2015). *World Development Report 2015: Mind, Society, and Behavior*. International Bank for Reconstruction and Development.

World Bank (2017). Propuesta de plan nacional de infraestructura educativa en Peru al 2025 : Metodologia de costeo de las intervenciones (Spanish). Technical report, World Bank Group.

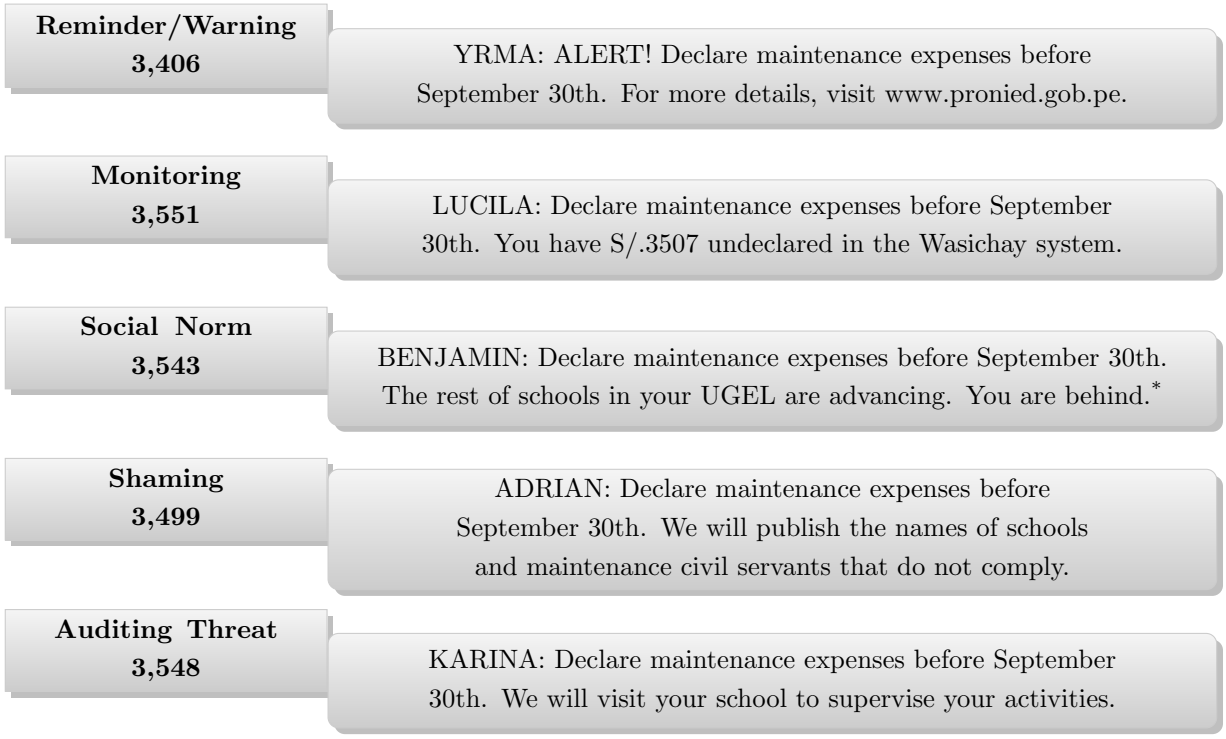
Young, H. (2015). The evolution of social norms. *Annual Review of Economics 7*, 359–387.

Figure 1: Intervention Cycle for PRONIED Maintenance Program in 2015



Note: Authors' construction based on administrative reports. Each circle represents a step during the intervention cycle. Maintenance CS stands for maintenance civil servant. UGEL infrastructure specialist is the official in charge of overseeing compliance with maintenance activities at the school district level. PRONIED specialist is the official in charge of overseeing compliance with maintenance activities at the national level. Specific activities during a particular step are described in bullet points.

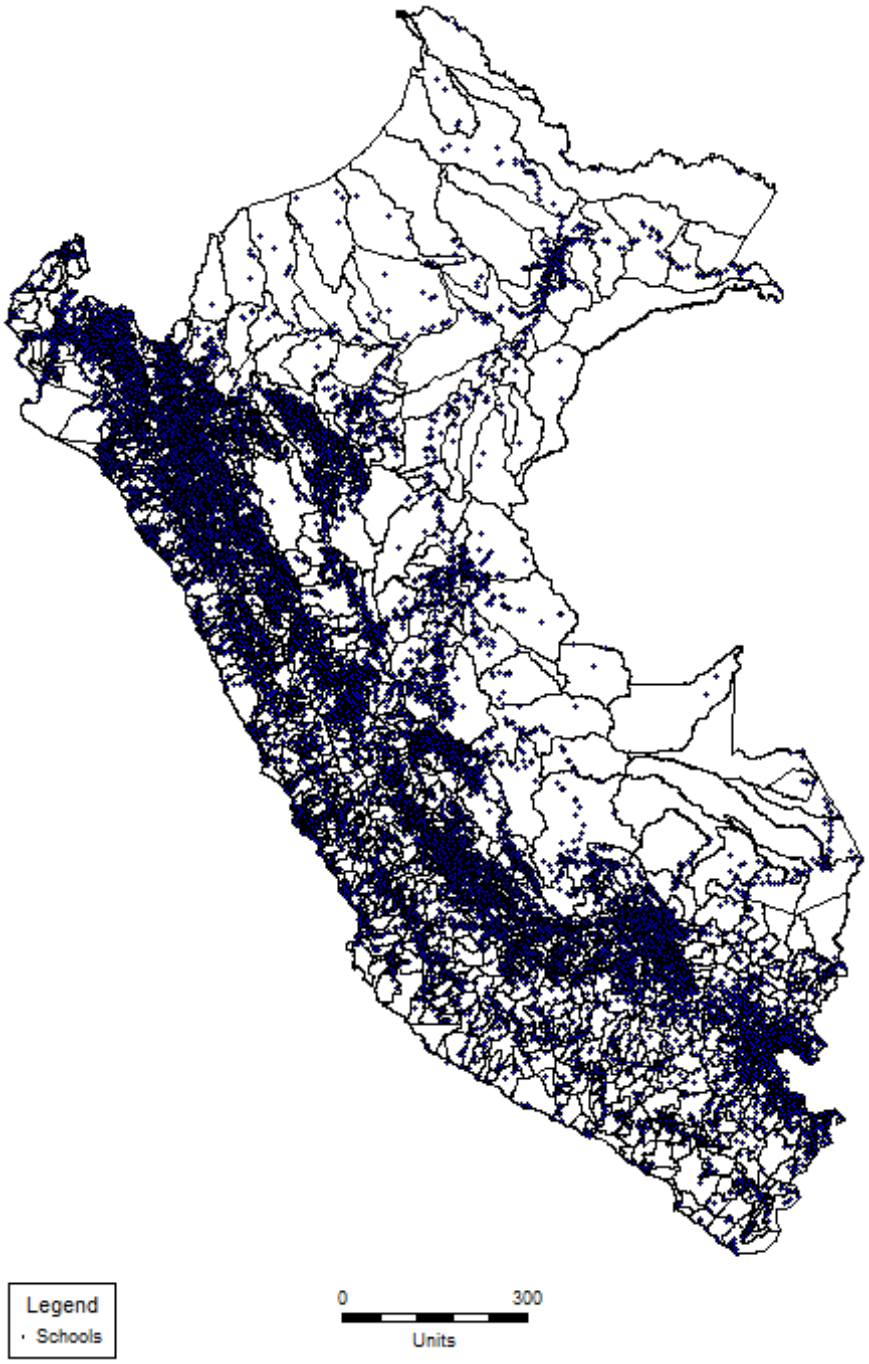
Figure 2: SMS Content in Benchmark Experiment in 2015



Note: Authors' elaboration. Each message includes the person's name and the deadline to comply with the activity. The rest of the content varies according to the behavioral principle to be emphasized. This example corresponds to the 3rd message delivered during the SMS campaign. All of the messages delivered are described in the Online Appendix (Table S2).

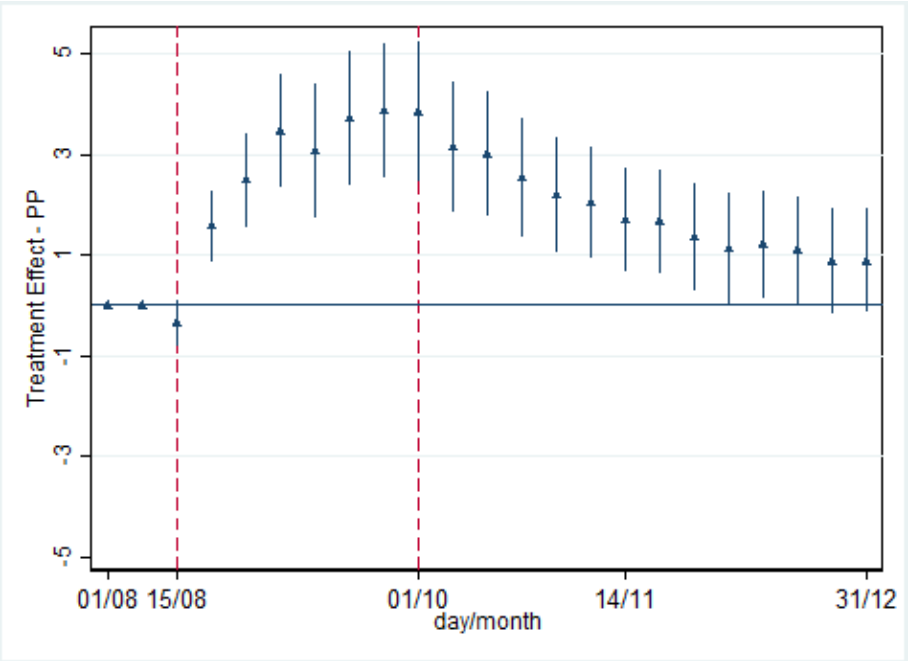
* Direct translation of the final sentence is difficult. The Spanish phrase is "falta usted," which literally means "you are missing."

Figure 3: Universe of Participating Schools in the Benchmark Experiment



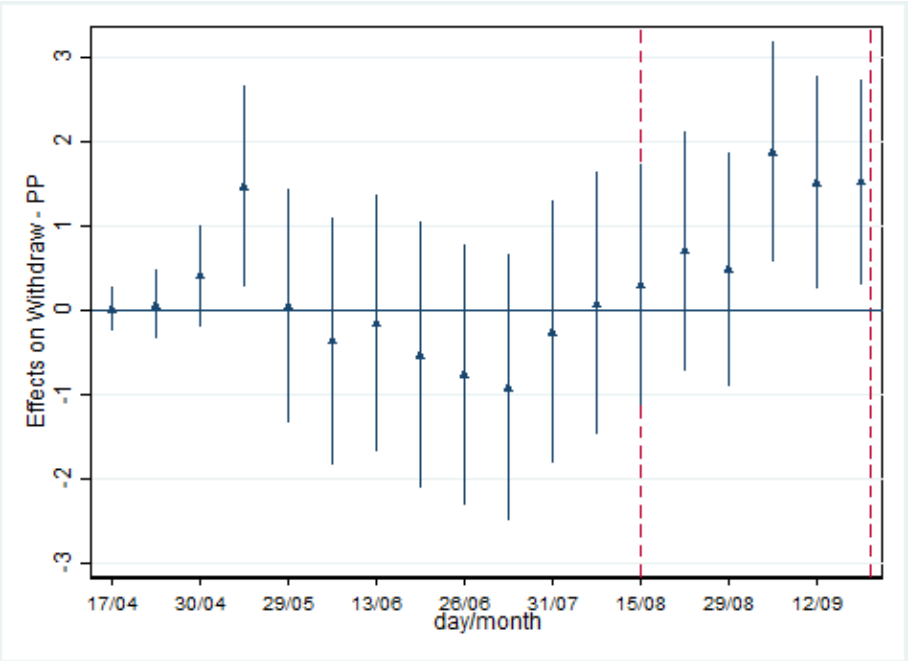
Note: Authors' elaboration. Each blue dot represents a school in the Benchmark Experiment in 2015.

Figure 4: Treatment effect on expense report submission, by week (Benchmark Experiment)



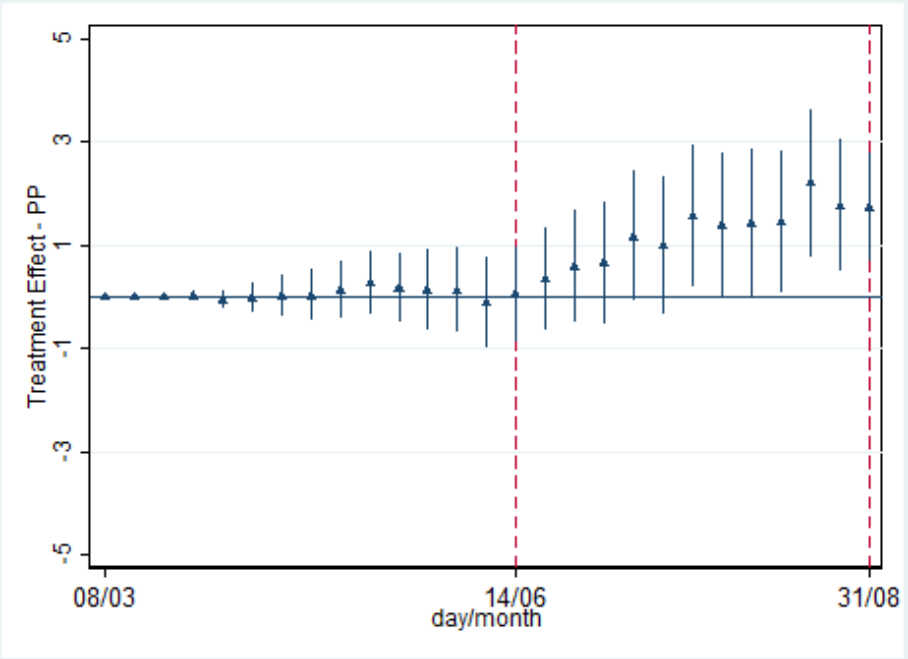
Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Figure 5: Treatment effect on withdrawal of 99% of bank balance, by week (Benchmark Experiment)



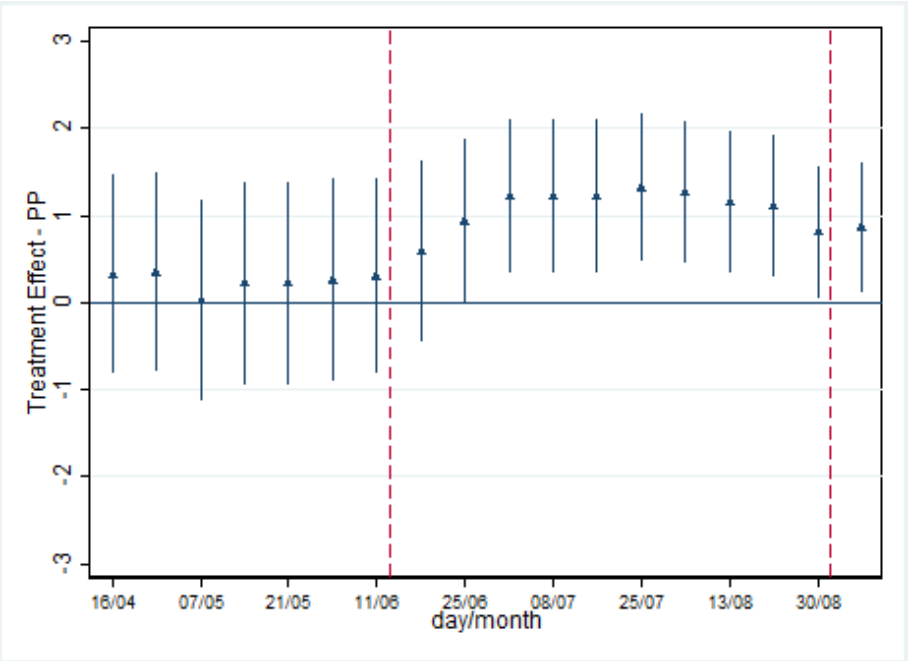
Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Figure 6: Treatment effect on expense report submission, by week (Follow-up Experiment)



Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Figure 7: Treatment effect on withdrawal of 99% of bank balance, by week (Follow-up Experiment)



Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Table 1. Descriptive Statistics for Benchmark Experiment in 2015

Variables	Mean	Std. Dev.	Min	Max	N
Panel A: Pre-treatment Outcomes					
Submitted to Maintenance Committee at Week 20	0.859	0.348	0	1	24,268
Submitted to Oversight Committee at Week 20	0.857	0.350	0	1	24,268
Submitted Technical Form at Week 20	0.707	0.455	0	1	24,268
Submitted Commitment Act at Week 20	0.674	0.469	0	1	24,268
Submitted Expense Report at Week 20	0.000	0.000	0	0	24,268
Submitted Oversight Report at Week 20	0.000	0.000	0	0	24,268
With Approved Expense Report at Week 20	0.000	0.000	0	0	24,268
Bank Balance at 26/06/2016	2,685	5,321	0.010	52,922	20,899
Panel B: Outcomes					
Submitted Commitment Act at Week 30	0.839	0.367	0	1	24,268
Submitted Expense Report at Week 30	0.769	0.421	0	1	24,268
Submitted Oversight Report at Week 30	0.102	0.302	0	1	24,268
With Approved Expense Report at Week 30	0.277	0.448	0	1	24,268
Withdrew Something	0.997	0.052	0	1	21,023
Withdrew 50%	0.929	0.256	0	1	21,023
Withdrew 80%	0.918	0.275	0	1	21,023
Withdrew 95%	0.905	0.293	0	1	21,023
Withdrew 99%	0.898	0.303	0	1	21,023
Panel C: Maintenance CS Characteristics					
Sex (% Men)	0.455	0.498	0	1	24,268
Age	46.1	8.1	20.9	70.3	24,268
Appointed Maintenance CS	0.276	0.447	0	1	24,268
Hired Maintenance CS	0.108	0.311	0	1	24,268
Allocation Transfer	7,733	7,972	0	30,000	24,268
Panel D: School Characteristics					
Classrooms	5.0	6.2	0	72	24,268
Students	128.1	592.7	0	77,990	24,268
Bathroom Connected to Public Drainage System	0.379	0.485	0	1	24,268
Bathroom Connected to Septic Tank	0.240	0.427	0	1	24,268
Bathroom Connected to a Black Well	0.262	0.440	0	1	24,268
Bathroom Connected to River, Ditch or Canal	0.028	0.165	0	1	24,268
No Bathroom	0.065	0.247	0	1	24,268
Total Land Area	6,746	37,833	0	1,000,000	24,268
Fully Fenced	0.327	0.469	0	1	24,268
Partially Fenced	0.251	0.434	0	1	24,268
Not Fenced	0.396	0.489	0	1	24,268
Number Educ-Admin Spaces	8.1	9.4	0	138	24,268
Number of Buildings	2.2	2.3	0	43	24,268
Average Leaks in Pavilions	1.2	1.5	0	41	24,268
Average Leaks	1.1	1.4	0	35	24,268
Panel E: District Characteristics					
Altitude	1,639	1,557	0	5,043	24,268
Area (% Rural)	0.603	0.489	0	1	19,365
Electricity	0.836	0.370	0	1	23,650
Public Drinking Water Network	0.669	0.470	0	1	23,650
Public Drainage Network	0.398	0.489	0	1	23,646
Internet Cafe	0.245	0.430	0	1	23,651
Bank Branch	0.123	0.328	0	1	23,651

Note: Authors' elaboration based on MINEDU's administrative records. The table reports the means, standard deviations, minimum and maximum values, and the sample size. Sample includes all maintenance civil servants who had not submitted their expense report at the beginning of the SMS campaign.

Table 2. Randomization Balance Analysis for Benchmark Experiment in 2015

Variables	Control	Reminder/ Warning	Social Norm	Monitoring	Shaming	Auditing threat	Joint Hyp. p-val.
Panel A: Pre-treatment Outcomes							
Submitted to Maintenance Committee at Week 20	0.862	0.859	0.855	0.855	0.860	0.859	0.884
	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	
Submitted to Oversight Committee at Week 20	0.860	0.858	0.853	0.852	0.858	0.858	0.861
	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	
Submitted Technical Form at Week 20	0.717	0.697	0.698	0.700	0.715	0.703	0.122
	(0.005)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	
Submitted Commitment Act at Week 20	0.688	0.666	0.666	0.664	0.677	0.671	0.071
	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	
Submitted Expense Report at Week 20	0.000	0.000	0.000	0.000	0.000	0.000	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Submitted Oversight Report to Week 20	0.000	0.000	0.000	0.000	0.000	0.000	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
With Approved Expense Report at Week 20	0.000	0.000	0.000	0.000	0.000	0.000	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Bank Balance at 26/06	2600.5	2704.7	2714.4	2764.4	2688.7	2714.5	0.795
	(67.448)	(99.288)	(96.709)	(99.677)	(97.203)	(97.595)	
Panel B: Maintenance CS Characteristics							
Sex (% Men)	0.459	0.46	0.448	0.456	0.445	0.462	0.606
	(0.006)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	
Age	46.1	46.2	46.1	46.1	46.2	46.4	0.344
	(0.099)	(0.139)	(0.138)	(0.135)	(0.139)	(0.133)	
Appointed Maintenance CS	0.276	0.284	0.274	0.277	0.264	0.285	0.438
	(0.005)	(0.008)	(0.007)	(0.008)	(0.007)	(0.008)	
Hired Maintenance CS	0.109	0.108	0.111	0.109	0.108	0.104	0.968
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	
Allocation Transfer	7751.1	7850.8	7566.2	7663.7	7697.7	7858.1	0.617
	(97.008)	(138.838)	(131.899)	(133.225)	(134.130)	(135.477)	
Panel C: School Characteristics							
Classrooms	5.0	5.1	4.8	4.9	4.9	5.1	0.181
	(0.076)	(0.107)	(0.100)	(0.099)	(0.102)	(0.109)	
Students	121.2	134.3	118.5	141.8	122.2	137.1	0.385
	(2.901)	(8.367)	(5.233)	(22.287)	(4.262)	(6.320)	
Bathroom Connected to Public Drainage System	0.38	0.378	0.37	0.382	0.387	0.374	0.744
	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	
Bathroom Connected to Septic Tank	0.243	0.244	0.235	0.239	0.231	0.247	0.560
	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	
Bathroom Connected to a Black Well	0.26	0.267	0.27	0.26	0.264	0.256	0.799
	(0.005)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	
Bathroom Connected to River, Ditch or Canal	0.029	0.027	0.029	0.026	0.028	0.028	0.983
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	
No Bathroom	0.064	0.059	0.069	0.068	0.066	0.067	0.556
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	
Total Land Area	7103.5	7494.5	5820.3	5991.9	6713.3	7059.2	0.335
	(453.294)	(743.779)	(448.382)	(573.630)	(687.226)	(723.130)	
Fully Fenced	0.329	0.33	0.314	0.323	0.339	0.322	0.298
	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	
Partially Fenced	0.249	0.25	0.264	0.253	0.244	0.25	0.465
	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	
Not Fenced	0.398	0.394	0.394	0.398	0.392	0.4	0.986
	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	
Number Educ-Admin Spaces	8.145	8.243	7.902	8.016	8.1	8.186	0.688
	(0.114)	(0.162)	(0.154)	(0.156)	(0.161)	(0.164)	
Number of Buildings	2.249	2.206	2.172	2.159	2.213	2.23	0.421
	(0.030)	(0.039)	(0.038)	(0.036)	(0.035)	(0.039)	
Average Leaks in Pavilions	1.17	1.181	1.178	1.147	1.142	1.154	0.822
	(0.018)	(0.028)	(0.026)	(0.025)	(0.024)	(0.026)	
Average Leaks	1.068	1.081	1.082	1.06	1.031	1.051	0.646
	(0.017)	(0.025)	(0.024)	(0.023)	(0.022)	(0.026)	

Table 2. Randomization Balance Analysis for Benchmark Experiment in 2015 (continued)

Variables	Control	Reminder/ Warning	Social Norm	Monitoring	Shaming	Auditing threat	Joint Hypothesis
Panel D: District Characteristics							
Altitude	1620.8	1658.7	1631.7	1672.8	1682.1	1587.4	0.081
	(18.874)	(26.750)	(26.243)	(26.218)	(26.426)	(26.063)	
Area (% Rural)	0.598	0.603	0.601	0.609	0.606	0.605	0.960
	(0.007)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	
Electricity	0.842	0.833	0.836	0.835	0.833	0.831	0.749
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	
Public Drinking Water Network	0.674	0.661	0.666	0.668	0.673	0.67	0.840
	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	
Public Drainage Network	0.398	0.392	0.391	0.401	0.405	0.396	0.832
	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	
Internet Cafe	0.25	0.242	0.24	0.244	0.246	0.242	0.909
	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	
Bank Branch	0.129	0.121	0.115	0.127	0.115	0.123	0.268
	(0.004)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	
Observations	6,723	3,406	3,543	3,551	3,499	3,546	

Note: Authors' elaboration. Sample includes all maintenance civil servants who had not submitted their expense report at the beginning of the SMS campaign. For each treatment arm, means and standard errors are reported for each pre-treatment variable. Final column is the p-value for the test of equality of means across all groups. Table S1 in the Online Appendix contains the variables' full definitions.

Table 3. Effect of SMS Campaign on WASICHAY Outcomes in the Benchmark Experiment

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	**Expense report**	**Oversight report**	**Approved expense report**	**Expense report**	**Oversight report**	**Approved expense report**
SMS	3.855***	0.765*	1.602**	4.160***	0.732*	1.677***
	(0.710)	(0.440)	(0.650)	(0.684)	(0.427)	(0.622)
Reminder/Warning	3.029***	1.165*	1.837*	3.179***	1.104	1.838*
	(0.856)	(0.693)	(0.975)	(0.851)	(0.691)	(0.945)
Social Norm	4.266***	1.884***	3.326***	4.675***	1.857***	3.373***
	(1.053)	(0.675)	(0.975)	(1.016)	(0.666)	(0.944)
Monitoring	4.862***	-0.0638	0.96	5.115***	-0.131	1.014
	(0.975)	(0.479)	(0.898)	(0.931)	(0.459)	(0.871)
Shaming	3.601***	0.395	1.005	3.771***	0.252	0.979
	(0.813)	(0.626)	(0.874)	(0.812)	(0.596)	(0.849)
Auditing threat	3.479***	0.457	0.886	4.016***	0.587	1.180
	(1.049)	(0.584)	(0.884)	(0.970)	(0.576)	(0.839)
Control mean	74.15	9.624	26.54	74.15	9.624	26.54
Controls	No	No	No	Yes	Yes	Yes
Observations	24,257	24,257	24,257	24,257	24,257	24,257

Note: Treatment effects and means are reported in percentage points. "SMS" pools all treatment arms. Rows below are from specifications that estimate separate effects for each treatment arm. Columns 4 to 6 include controls for personal characteristics (age, gender, type of contract), school characteristics (number of classrooms, number of buildings, land area, number of students, bathroom characteristics, distance to UGEL) and municipality characteristics (altitude, access to electricity, access to drinking water network, access to internet, availability of a bank branch), including dummies for missing observations. Table S1 in the Online Appendix contains the outcome, treatment and control variables' full definitions. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% respectively.

Table 4. Effect of SMS Campaign on Withdrawal of Maintenance Funds in the Benchmark Experiment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Variables	**Withdraw Something**	**Withdraw 50%**	**Withdraw 80%**	**Withdraw 95%**	**Withdraw 99%**	**Withdraw Something**	**Withdraw 50%**	**Withdraw 80%**	**Withdraw 95%**	**Withdraw 99%**
SMS	0.043	0.534	0.916*	1.048**	1.458**	0.057	0.664	1.028**	1.132**	1.527***
	(0.081)	(0.463)	(0.485)	(0.523)	(0.572)	(0.076)	(0.409)	(0.437)	(0.481)	(0.531)
Reminder/Warning	0.071	0.360	0.797	1.030	1.445**	0.086	0.424	0.867	1.096*	1.511**
	(0.100)	(0.595)	(0.629)	(0.660)	(0.715)	(0.100)	(0.547)	(0.587)	(0.631)	(0.682)
Social Norm	0.111	0.160	0.465	0.569	1.036	0.142	0.475	0.742	0.773	1.204*
	(0.126)	(0.651)	(0.682)	(0.726)	(0.766)	(0.121)	(0.575)	(0.605)	(0.661)	(0.711)
Monitoring	0.046	0.952*	1.389**	1.624***	2.125***	0.057	1.061**	1.479***	1.683***	2.172***
	(0.135)	(0.563)	(0.558)	(0.607)	(0.660)	(0.133)	(0.524)	(0.522)	(0.576)	(0.627)
Shaming	-0.058	0.647	1.067*	1.152*	1.644**	-0.051	0.706	1.105*	1.168*	1.648**
	(0.128)	(0.594)	(0.621)	(0.677)	(0.726)	(0.123)	(0.558)	(0.586)	(0.639)	(0.686)
Auditing threat	0.046	0.547	0.857	0.867	1.044	0.050	0.648	0.944	0.941	1.101
	(0.102)	(0.622)	(0.670)	(0.709)	(0.762)	(0.0959)	(0.566)	(0.620)	(0.661)	(0.715)
Control mean	99.693	92.546	91.097	89.783	88.743	99.693	92.546	91.097	89.783	88.743
Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	21,012	21,012	21,012	21,012	21,012	21,012	21,012	21,012	21,012	21,012

Note: Treatment effects and means are reported in percentage points. "SMS" pools all treatment arms. Rows below are from specifications that estimate separate effects for each treatment arm. Columns 6 to 10 include controls for personal characteristics (age, gender, type of contract), school characteristics (number of classrooms, number of buildings, land area, number of students, bathroom characteristics, distance to UGEL) and municipality characteristics (altitude, access to electricity, access to drinking water network, access to internet, availability of a bank branch), including dummies for missing observations. Table S1 in the Online Appendix contains the outcome, treatment and control variables' full definitions. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% respectively.

Table 5. Effects of Behavioral-based SMS Campaign on WASICHAY Outcomes in the Follow-Up Experiment

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	**Expense report**	**Oversight report**	**Approved expense report**	**Expense report**	**Oversight report**	**Approved expense report**
SMS	1.733***	1.084**	1.240*	1.709***	1.024*	1.204*
	(0.536)	(0.528)	(0.674)	(0.536)	(0.527)	(0.674)
Descriptive Social Norm	1.752***	1.154*	1.211	1.776***	1.088*	1.184
	(0.568)	(0.586)	(0.749)	(0.561)	(0.581)	(0.741)
Injunctive Social Norm	1.683**	0.986	1.680*	1.546**	0.864	1.522*
	(0.701)	(0.715)	(0.873)	(0.703)	(0.719)	(0.871)
Social Benefit	1.743***	1.056*	0.986	1.728***	1.045*	1.020
	(0.658)	(0.614)	(0.790)	(0.658)	(0.612)	(0.788)
Control mean	80.62	17.37	39.19	80.62	17.37	39.19
Controls	No	No	No	Yes	Yes	Yes
Observations	31,947	31,947	31,947	31,947	31,947	31,947

Note: Treatment effects and means are reported in percentage points. "SMS" pools all treatment arms. Rows below are from specifications that estimate separate effects for each treatment arm. Columns 4 to 6 include controls for personal characteristics (age, gender, type of contract), school characteristics (number of classrooms, number of buildings, land area, number of students, bathroom characteristics, distance to UGEL) and municipality characteristics (altitude, access to electricity, access to drinking water network, access to internet, availability of a bank branch), including dummies for missing observations. Table S1 in the Online Appendix contains the outcome, treatment and control variables' full definitions. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% respectively.

Table 6. Effect of SMS Campaign on Withdrawal of Maintenance Funds in the Follow-Up Experiment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Variables	**Withdraw Something**	**Withdraw 50%**	**Withdraw 80%**	**Withdraw 95%**	**Withdraw 99%**	**Withdraw Something**	**Withdraw 50%**	**Withdraw 80%**	**Withdraw 95%**	**Withdraw 99%**
SMS	0.320	0.400	0.396	0.697	0.859*	0.353	0.436	0.434	0.736*	0.901**
	(0.287)	(0.357)	(0.403)	(0.444)	(0.453)	(0.283)	(0.353)	(0.400)	(0.441)	(0.449)
Descriptive Social Norm	0.293	0.449	0.551	0.865*	0.981*	0.323	0.483	0.582	0.891*	1.011**
	(0.325)	(0.408)	(0.448)	(0.490)	(0.502)	(0.321)	(0.400)	(0.440)	(0.480)	(0.491)
Injunctive Social Norm	0.461	0.626	0.609	0.948*	1.098**	0.519	0.695	0.680	1.018*	1.171**
	(0.366)	(0.423)	(0.489)	(0.525)	(0.533)	(0.363)	(0.425)	(0.491)	(0.530)	(0.536)
Social Benefit	0.263	0.183	0.0479	0.307	0.536	0.282	0.200	0.0731	0.340	0.574
	(0.312)	(0.405)	(0.468)	(0.512)	(0.519)	(0.304)	(0.401)	(0.468)	(0.511)	(0.518)
Control mean	96.242	93.424	91.732	90.380	89.891	96.242	93.424	91.732	90.380	89.891
Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	31,938	31,938	31,938	31,938	31,938	31,938	31,938	31,938	31,938	31,938

Note: Treatment effects and means are reported in percentage points. "SMS" pools all treatment arms. Rows below are from specifications that estimate separate effects for each treatment arm. Columns 6 to 10 include controls for personal characteristics (age, gender, type of contract), school characteristics (number of classrooms, number of buildings, land area, number of students, bathroom characteristics, distance to UGEL) and municipality characteristics (altitude, access to electricity, access to drinking water network, access to internet, availability of a bank branch), including dummies for missing observations. Table S1 in the Online Appendix contains the outcome, treatment and control variables' full definitions. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% respectively.

Table 7. Persistence of Treatment Effects between 2015 and 2016

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	**Expense report**	**Oversight report**	**Approved Expense report**	**Expense report**	**Oversight report**	**Approved Expense report**
			Panel A. Control 2015 - Treated 2016			
Control - Treated	2.323*	0.769	0.707	2.155*	0.616	0.546
	(1.191)	(1.185)	(1.672)	(1.180)	(1.190)	(1.630)
Observations	6,642	6,642	6,642	6,642	6,642	6,642
			Panel B. Treated both years			
Treated - Treated	2.507**	0.830	0.883	2.502**	0.565	0.857
	(1.111)	(1.071)	(1.517)	(1.104)	(1.066)	(1.493)
Observations	15,489	15,489	15,489	15,489	15,489	15,489
			Panel C. Treated 2015 - Control 2016			
Treated - Control	-0.139	-0.311	-1.503	-0.191	-0.384	-1.377
	(1.411)	(1.351)	(1.771)	(1.396)	(1.386)	(1.749)
Observations	4,017	4,017	4,017	4,017	4,017	4,017
Control mean	80.62	17.37	39.19	80.62	17.37	39.19
Controls	No	No	No	Yes	Yes	Yes

Note: Panel A shows estimated effects of 2016 SMS campaign for civil servants who were in the control group in 2015, compared to the never-treated group. Panel B shows estimated effects in 2016 for civil servants who were treated in both periods, compared to the never-treated group. Panel C shows the estimated "effects" of the 2016 campaign for those who were treated in 2015 and control in 2016, compared to the never-treated group. Robust standard errors clustered at the UGEL in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% respectively.

Table 8. External Validity: SMS Effects in CUNA MAS Experiment

Variables	(1)	(2)	(3)	(4)
	Compliance	Compliance	Compliance	Compliance
SMS	2.840	3.016	3.440*	4.446**
	(2.261)	(2.112)	(1.908)	(1.905)
Social Norm	0.721	1.167	1.269	2.352
	(2.717)	(2.515)	(2.286)	(2.318)
Monitoring	4.828*	4.741**	5.471***	6.404***
	(2.523)	(2.379)	(2.108)	(2.070)
Control mean	70.14	70.14	70.14	70.14
Observations	5,373	5,373	5,368	3,220
Controls	No	Yes	Yes	Yes
Baseline	No	No	Yes	Yes
October&December	Yes	Yes	Yes	No

Note: Compliance is the percentage of home visits for which the CUNA MAS civil servant has submitted an online report. The program target is 100%. "SMS" pools both treatment arms. Rows below are from specifications that estimate separate effects for each treatment arm. Columns 3 and 4 include controls for gender, length of tenure on the job, and fixed effects for the regional office (Territorial Unit) overseeing the civil servant. Standard errors clustered at the civil servant level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% respectively.

Table 9. Difference Between Expense Report and Technical Form

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Variables	**Total**	**Ceilings**	**Sanitary Facilities**	**Floors**	**Walls**	**Doors**	**Windows**	**Electrical Installations**	**Furniture Repairs**	**Paint**	**Furniture Replacement**	**School Supplies**
						Panel A. Without controls						
SMS	32.653***	6.149	-6.992	6.630	-8.507	-1.566	4.332	1.288	-1.386	11.871	14.239	5.149*
	(10.007)	(7.657)	(9.122)	(8.207)	(8.186)	(5.493)	(3.844)	(4.113)	(8.584)	(7.680)	(10.022)	(2.944)
Descriptive Social Norm	34.347***	5.475	-2.843	3.191	-10.963	-2.291	3.287	1.743	-3.066	14.583	17.306	4.853
	(9.874)	(8.517)	(9.766)	(9.089)	(8.847)	(5.886)	(4.258)	(4.445)	(9.447)	(8.859)	(10.797)	(3.240)
Injunctive Social Norm	24.501*	7.582	-12.487	6.834	-1.347	-4.075	6.605	1.168	-8.003	6.808	18.088	5.085
	(13.355)	(9.061)	(10.460)	(10.133)	(10.669)	(6.473)	(4.639)	(5.645)	(9.783)	(9.867)	(12.679)	(3.510)
Social Benefit	35.818***	6.091	-8.853	11.064	-10.003	1.066	4.209	0.763	5.245	11.632	7.605	5.585*
	(11.492)	(9.389)	(10.059)	(9.486)	(9.203)	(6.259)	(4.499)	(4.613)	(9.302)	(8.465)	(10.627)	(3.302)
						Panel B. With controls						
SMS	32.132***	5.950	-7.702	6.685	-8.588	-1.496	4.307	0.978	-1.217	12.077	14.696	5.173*
	(9.905)	(7.568)	(9.143)	(8.220)	(8.212)	(5.511)	(3.860)	(4.131)	(8.583)	(7.713)	(10.035)	(2.950)
Descriptive Social Norm	33.965***	5.331	-3.743	3.415	-11.029	-2.107	3.246	1.360	-3.072	14.950*	17.725	4.932
	(9.694)	(8.374)	(9.811)	(9.110)	(8.870)	(5.894)	(4.283)	(4.470)	(9.447)	(8.929)	(10.803)	(3.239)
Injunctive Social Norm	24.017*	7.049	-12.520	6.881	-1.524	-4.062	6.677	0.894	-7.781	7.017	18.475	4.961
	(13.277)	(9.089)	(10.350)	(10.187)	(10.708)	(6.503)	(4.662)	(5.657)	(9.789)	(9.876)	(12.751)	(3.516)
Social Benefit	35.086***	6.043	-9.765	10.904	-10.036	1.021	4.144	0.524	5.612	11.619	8.157	5.635*
	(11.524)	(9.272)	(10.142)	(9.456)	(9.201)	(6.288)	(4.495)	(4.632)	(9.262)	(8.466)	(10.639)	(3.314)
Control mean	-77.37	-29.71	-6.920	-1.947	19.46	-7.057	-17.34	-2.744	-10.09	17.87	-19.80	-12.74
Observations	28,171	28,171	28,171	28,171	28,171	28,171	28,171	28,171	28,171	28,171	28,171	28,171

Note: Dependent variables are difference between the expenditure amount declared in the expense report and the amount dedicated to that area in the technical form. "SMS" pools all treatment arms. Rows below are from specifications that estimate separate effects for each treatment arm. Panel B include controls for personal characteristics (age, gender, type of contract), school characteristics (number of classrooms, number of buildings, land area, number of students, bathroom characteristics, distance to UGEL) and municipality characteristics (altitude, access to electricity, access to drinking water network, access to internet, availability of a bank branch), including dummies for missing observations. Table S1 in the Online Appendix contains the outcome, treatment and control variables' full definitions. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% respectively.

Table 10. Effects on Infrastructure Quality

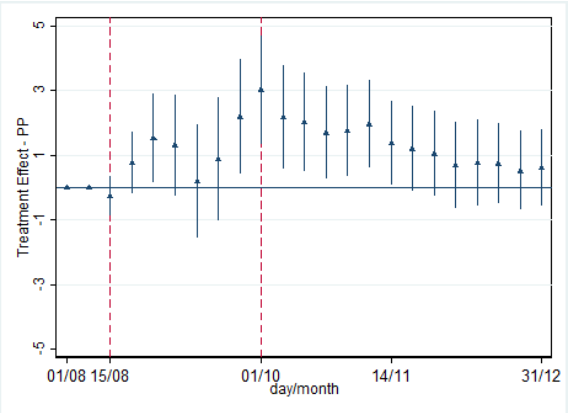
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Variables	**Walls**	**Ceilings**	**Floors**	**Windows**	**Doors**	**Bathroom Sinks**	**Walls**	**Ceilings**	**Floors**	**Windows**	**Doors**	**Bathroom Sinks**
				in bad condition						**in bad condition**		
SMS	0.102	-0.278	0.193	-0.686	-1.134	-2.902*	0.047	-0.320	0.171	-0.693	-1.145	-3.076**
	(0.572)	(0.690)	(0.874)	(0.535)	(0.933)	(1.565)	(0.565)	(0.682)	(0.873)	(0.545)	(0.912)	(1.524)
Descriptive Social Norm	-0.099	-0.149	-0.279	-0.930	-1.602	-2.135	-0.151	-0.215	-0.319	-0.983	-1.577	-2.457
	(0.649)	(0.763)	(0.991)	(0.583)	(1.019)	(1.783)	(0.633)	(0.749)	(0.978)	(0.595)	(1.008)	(1.769)
Injunctive Social Norm	0.496	-0.228	1.359	-0.010	0.228	-1.887	0.440	-0.197	1.393	0.035	0.220	-1.890
	(0.854)	(0.844)	(1.198)	(0.685)	(1.184)	(2.002)	(0.856)	(0.845)	(1.202)	(0.672)	(1.154)	(1.969)
Social Benefit	0.109	-0.482	0.046	-0.811	-1.415	-4.563**	0.050	-0.540	0.009	-0.794	-1.479	-4.655***
	(0.635)	(0.746)	(1.031)	(0.626)	(1.033)	(1.853)	(0.635)	(0.742)	(1.039)	(0.644)	(1.016)	(1.736)
Control mean	4.13	4.62	8.09	4.70	9.59	58.98	4.13	4.62	8.09	4.70	9.59	58.98
Controls	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,351	7,359	7,359	7,359	7,359	6,650	7,343	7,351	7,351	7,351	7,351	6,650

Note: Dependent variables are dummies equal to 1 if the surveyed teacher considered that the quality of that item of infrastructure was bad. "SMS" pools all treatment arms. Rows below are from specifications that estimate separate effects for each treatment arm. Columns 7 to 12 include controls for personal characteristics (age, gender, type of contract), school characteristics (number of classrooms, number of buildings, land area, number of students, bathroom characteristics, distance to UGEL) and municipality characteristics (altitude, access to electricity, access to drinking water network, access to internet, availability of a bank branch), including dummies for missing observations. Table S1 in the Online Appendix contains the outcome, treatment and control variables' full definitions. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% respectively.

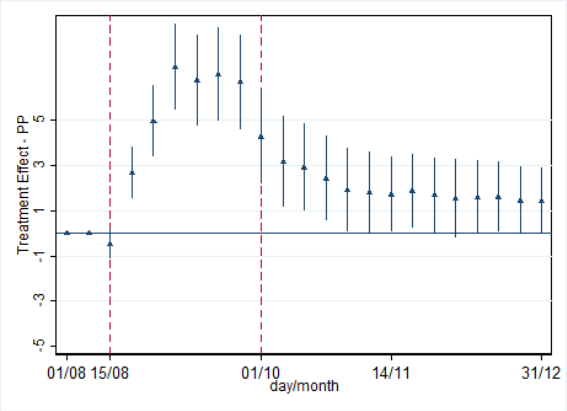
Appendix: Figures and Tables

Figure A1: Treatment effect on expense report submission, by week, separated by treatment arm (Benchmark Experiment)

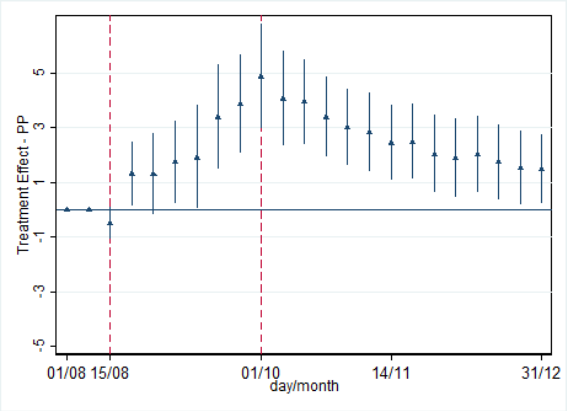
(a) Reminder/Warning



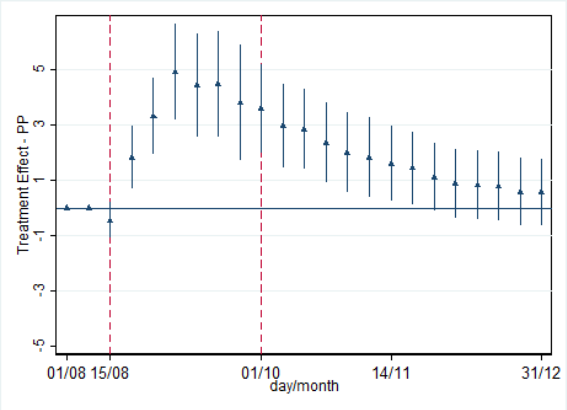
(b) Monitoring



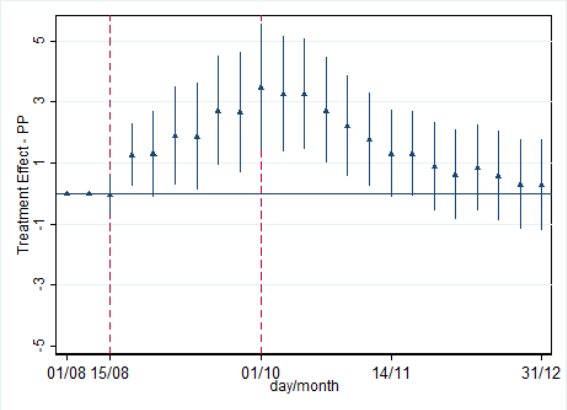
(c) Social Norm



(d) Shaming



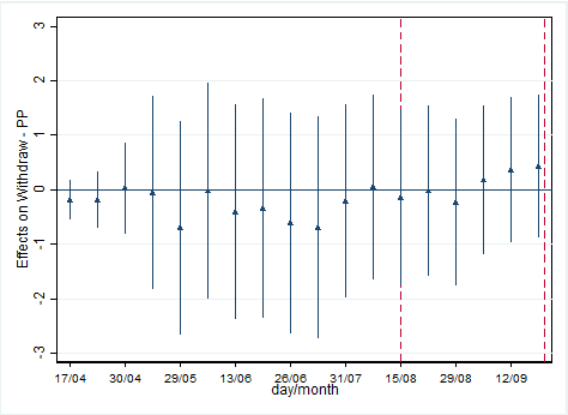
(e) Auditing Threat



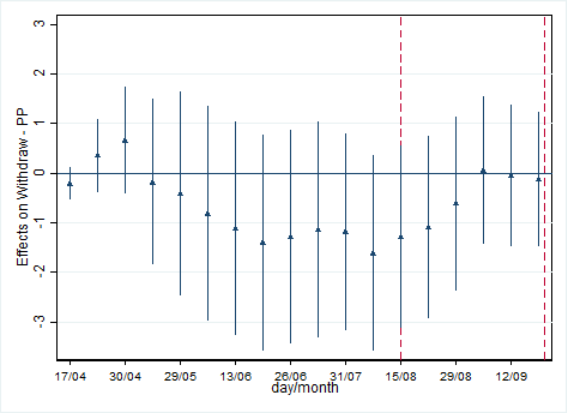
Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Figure A2: Treatment effect on withdrawal of 99% of bank balance, by week, separated by treatment arm (Benchmark Experiment)

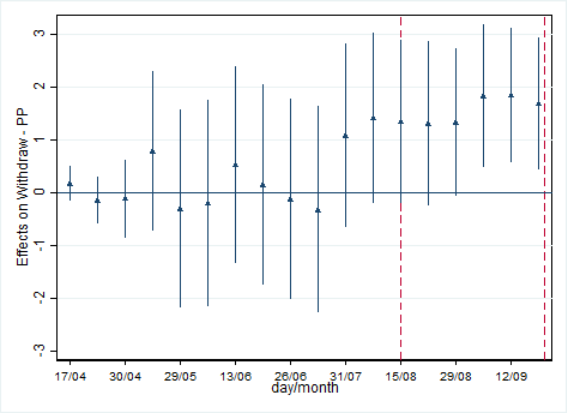
(a) Reminder/Warning



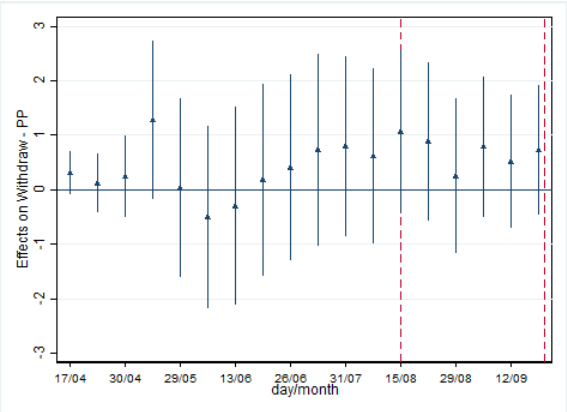
(b) Monitoring



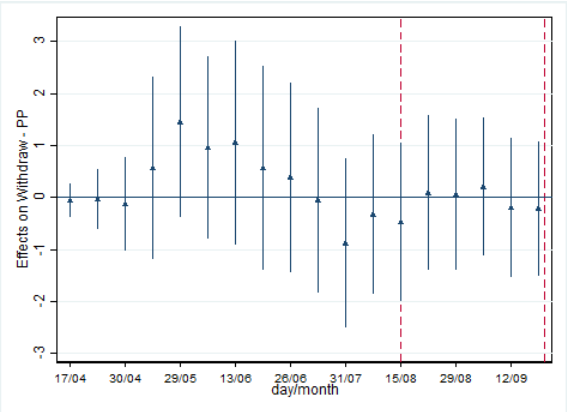
(c) Social Norm



(d) Shaming



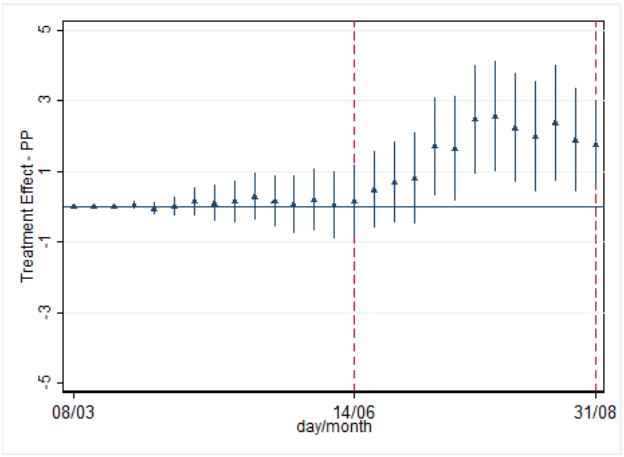
(e) Auditing Threat



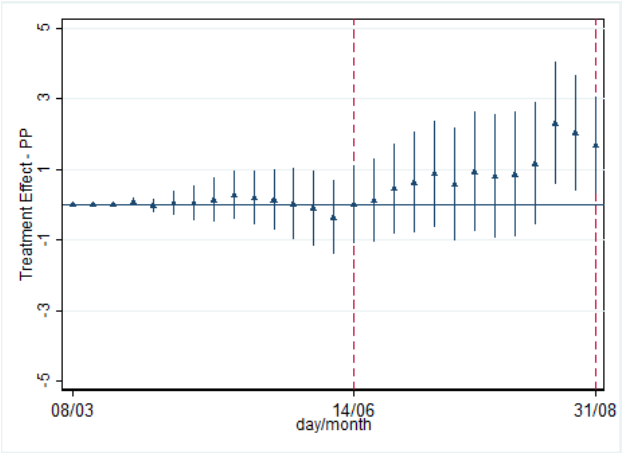
Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Figure A3: Treatment effect on expense report submission, by week, separated by treatment arm (Follow-up Experiment)

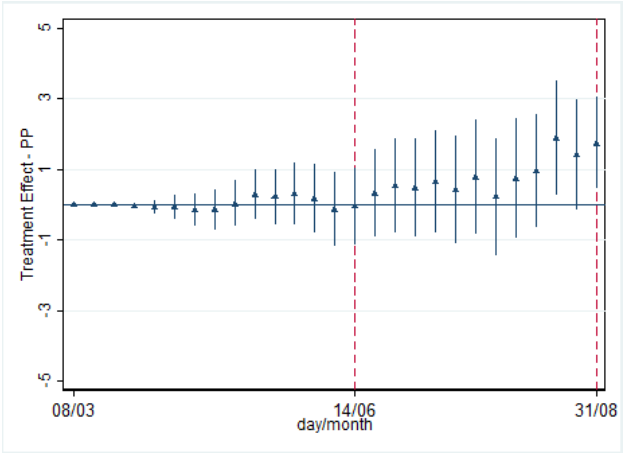
(a) Descriptive Social Norm



(b) Injunctive Social Norm



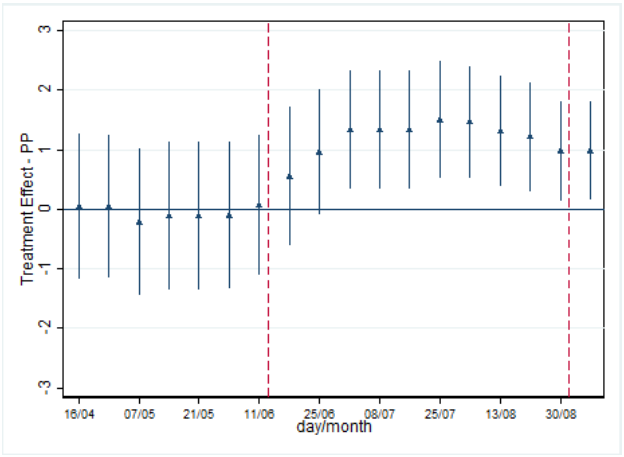
(c) Social Benefit



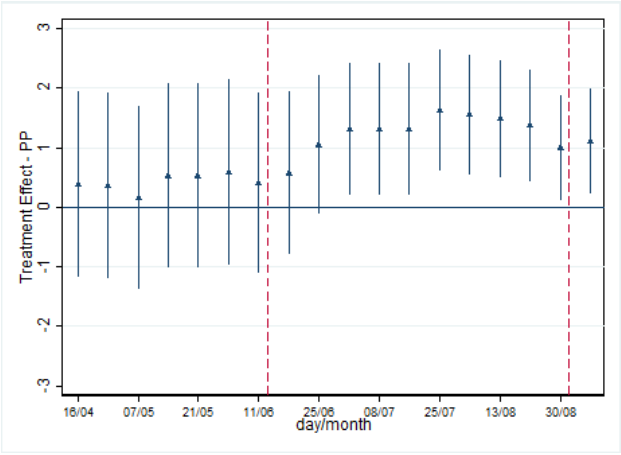
Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Figure A4: Treatment effect on withdrawal of 99% of bank balance, by week, separated by treatment arm (Follow-up Experiment)

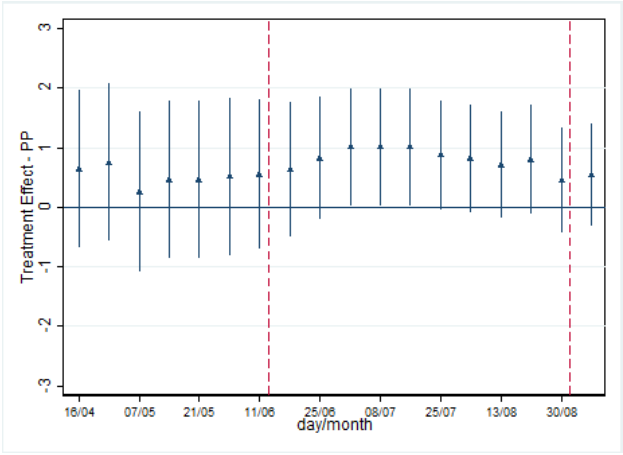
(a) Descriptive Social Norm



(b) Injunctive Social Norm



(c) Social Benefit



Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Table A1. Effect of SMS Campaign on WASICHAY Outcomes in the Follow-Up Experiment

	(1)	(2)	(3)	(4)	(5)	(6)
Variable	**Expense report**	**Oversight report**	**Approved Expense report**	**Expense report**	**Oversight report**	**Approved Expense report**
SMS	1.733***	1.084**	1.240*	1.709***	1.024*	1.204*
	(0.536)	(0.528)	(0.674)	(0.011)	(0.536)	(0.527)
Desc Norm Qual UGEL	2.573***	0.958	0.983	2.731***	0.951	0.998
	(0.850)	(0.817)	(1.036)	(0.843)	(0.809)	(1.013)
Desc Norm Quant UGEL	1.057	1.628*	1.612	1.094	1.598*	1.586
	(0.840)	(0.895)	(1.067)	(0.828)	(0.854)	(1.035)
Desc Norm Qual Peru	1.644**	0.839	0.368	1.666**	0.743	0.407
	(0.760)	(0.817)	(1.091)	(0.761)	(0.826)	(1.087)
Desc Norm Quant Peru	1.733**	1.189	1.883*	1.613**	1.063	1.743*
	(0.813)	(0.834)	(1.022)	(0.806)	(0.851)	(1.039)
Inj Norm Parents	1.131	0.541	0.922	0.950	0.390	0.698
	(0.830)	(0.889)	(1.128)	(0.831)	(0.895)	(1.112)
Inj Norm Principals	2.235***	1.432*	2.438**	2.142**	1.339*	2.348**
	(0.835)	(0.809)	(1.000)	(0.833)	(0.799)	(0.997)
Soc Ben Well-Being	0.375	-0.327	-0.403	0.303	-0.346	-0.391
	(0.951)	(0.850)	(1.131)	(0.944)	(0.847)	(1.118)
Soc Ben Pride	2.207**	1.662**	1.443	2.315***	1.686**	1.569
	(0.855)	(0.809)	(0.981)	(0.855)	(0.807)	(0.974)
Soc Ben Learning	2.646***	1.831**	1.916*	2.568***	1.794**	1.881*
	(0.797)	(0.890)	(1.059)	(0.783)	(0.889)	(1.065)
Control mean	80.62	17.37	39.19	80.62	17.37	39.19
Controls	No	No	No	Yes	Yes	Yes
Observations	31,947	31,947	31,947	31,947	31,947	31,947

Note: Treatment effects and means are reported in percentage points. "SMS" pools all treatment arms. Rows below are from specifications that estimate separate effects for each treatment arm. Columns 4 to 6 include controls for personal characteristics (age, gender, type of contract), school characteristics (number of classrooms, number of buildings, land area, number of students, bathroom characteristics, distance to UGEL) and municipality characteristics (altitude, access to electricity, access to drinking water network, access to internet, availability of a bank branch), including dummies for missing observations. Table S1 in the Online Appendix contains the outcome, treatment and control variables' full definitions. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% respectively.

Table A2. Effect of SMS Campaign on Withdrawal of Maintenance Funds in the Follow-Up Experiment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Variables	**Withdraw Something**	**Withdraw 50%**	**Withdraw 80%**	**Withdraw 95%**	**Withdraw 99%**	**Withdraw Something**	**Withdraw 50%**	**Withdraw 80%**	**Withdraw 95%**	**Withdraw 99%**
SMS	0.320	0.400	0.396	0.697	0.859*	0.353	0.436	0.434	0.736*	0.901**
	(0.287)	(0.357)	(0.403)	(0.444)	(0.453)	(0.283)	(0.353)	(0.400)	(0.441)	(0.449)
Desc Norm Qual UGEL	0.435	0.689	0.350	0.692	0.873	0.371	0.419	0.497	0.562	0.549
	(0.512)	(0.683)	(0.792)	(0.824)	(0.846)	(0.425)	(0.539)	(0.608)	(0.665)	(0.670)
Desc Norm Quant UGEL	0.439	0.865	1.035	1.310	1.626**	0.307	0.454	0.395	0.954	1.177*
	(0.476)	(0.671)	(0.787)	(0.813)	(0.806)	(0.453)	(0.600)	(0.633)	(0.660)	(0.659)
Desc Norm Qual Peru	0.544	1.277**	1.651**	1.690**	1.769**	0.367	0.720	0.941	1.226*	1.210*
	(0.474)	(0.615)	(0.692)	(0.742)	(0.766)	(0.404)	(0.532)	(0.603)	(0.655)	(0.685)
Desc Norm Quant Peru	0.709	0.966	1.136	1.513**	1.727**	0.248	0.343	0.495	0.822	1.107*
	(0.528)	(0.675)	(0.722)	(0.748)	(0.760)	(0.424)	(0.530)	(0.562)	(0.582)	(0.605)
Inj Norm Parents	0.474	0.767	1.106*	1.449**	1.495**	0.318	0.676	0.821	1.156**	1.276**
	(0.449)	(0.563)	(0.605)	(0.625)	(0.660)	(0.408)	(0.472)	(0.513)	(0.541)	(0.561)
Inj Norm Principals	0.976*	1.129*	1.535**	1.639**	1.752**	0.719*	0.713	0.540	0.881	1.066
	(0.524)	(0.652)	(0.711)	(0.765)	(0.769)	(0.434)	(0.551)	(0.644)	(0.694)	(0.690)
Soc Ben Well-Being	0.297	-0.160	0.0756	-0.0567	0.259	-0.0583	-0.480	-0.622	-0.456	-0.103
	(0.484)	(0.643)	(0.731)	(0.797)	(0.791)	(0.380)	(0.544)	(0.630)	(0.699)	(0.704)
Soc Ben Pride	0.301	-0.187	0.152	0.359	0.607	0.477	0.527	0.357	0.750	0.942
	(0.547)	(0.669)	(0.749)	(0.762)	(0.757)	(0.415)	(0.522)	(0.626)	(0.646)	(0.653)
Soc Ben Learning	0.842	1.403**	1.674***	1.576**	1.756***	0.429	0.555	0.485	0.728	0.882
	(0.521)	(0.594)	(0.617)	(0.647)	(0.668)	(0.418)	(0.547)	(0.571)	(0.619)	(0.641)
Control mean	96.242	93.424	91.732	90.380	89.891	96.242	93.424	91.732	90.380	89.891
Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	31,938	31,938	31,938	31,938	31,938	31,938	31,938	31,938	31,938	31,938

Note: Treatment effects and means are reported in percentage points. "SMS" pools all treatment arms. Rows below are from specifications that estimate separate effects for each treatment arm. Columns 6 to 10 include controls for personal characteristics (age, gender, type of contract), school characteristics (number of classrooms, number of buildings, land area, number of students, bathroom characteristics, distance to UGEL) and municipality characteristics (altitude, access to electricity, access to drinking water network, access to internet, availability of a bank branch), including dummies for missing observations. Table S1 in the Online Appendix contains the outcome, treatment and control variables' full definitions. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% respectively.

Table A3. Effect of Intensity of the SMS Campaign on WASICHAY Outcomes, Follow-Up Experiment

	(1)	(2)	(3)	(4)	(5)	(6)
Variable	**Expense report**	**Oversight report**	**Approved Expense report**	**Expense report**	**Oversight report**	**Approved Expense report**
SMS	1.733***	1.084**	1.240*	1.709***	1.024*	1.204*
	(0.536)	(0.528)	(0.674)	(0.536)	(0.527)	(0.674)
Long duration	1.532***	0.868	1.057	1.543***	0.792	1.002
	(0.584)	(0.565)	(0.719)	(0.568)	(0.567)	(0.715)
Short duration	1.935***	1.299**	1.424*	1.875***	1.256**	1.406*
	(0.582)	(0.561)	(0.726)	(0.593)	(0.555)	(0.725)
Control mean	80.62	17.37	39.19	80.62	17.37	39.19
Controls	No	No	No	Yes	Yes	Yes
Observations	31,947	31,947	31,947	31,947	31,947	31,947

Note: Treatment effects and means are reported in percentage points. "SMS" pools all treatment arms. Rows below are from specifications that estimate separate effects for each treatment arm. Columns 4 to 6 include controls for personal characteristics (age, gender, type of contract), school characteristics (number of classrooms, number of buildings, land area, number of students, bathroom characteristics, distance to UGEL) and municipality characteristics (altitude, access to electricity, access to drinking water network, access to internet, availability of a bank branch), including dummies for missing observations. Table S1 in the Online Appendix contains the outcome, treatment and control variables' full definitions. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% respectively.

Table A4. Effect of Intensity of the SMS Campaign on Withdrawal of Maintenance Funds, Follow-Up Experiment

	(1)	(2)	(3)	(4)	(5)	6)	(7)	(8)	(9)	(10)
Variables	**Withdraw Something**	**Withdraw 50%**	**Withdraw 80%**	**Withdraw 95%**	**Withdraw 99%**	**Withdraw Something**	**Withdraw 50%**	**Withdraw 80%**	**Withdraw 95%**	**Withdraw 99%**
SMS	0.320	0.400	0.396	0.697	0.859*	0.353	0.436	0.434	0.736*	0.901**
	(0.287)	(0.357)	(0.403)	(0.444)	(0.453)	(0.283)	(0.353)	(0.400)	(0.441)	(0.449)
Long duration	0.279	0.175	0.123	0.551	0.739	0.336	0.248	0.201	0.631	0.822*
	(0.292)	(0.378)	(0.435)	(0.482)	(0.488)	(0.287)	(0.373)	(0.429)	(0.475)	(0.480)
Short duration	0.361	0.625	0.670	0.843*	0.978**	0.370	0.624	0.667	0.841*	0.979**
	(0.314)	(0.386)	(0.428)	(0.468)	(0.478)	(0.311)	(0.381)	(0.427)	(0.467)	(0.478)
Control mean	96.242	93.424	91.732	90.380	89.891	96.242	93.424	91.732	90.380	89.891
Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	31,938	31,938	31,938	31,938	31,938	31,938	31,938	31,938	31,938	31,938

Note: Treatment effects and means are reported in percentage points. "SMS" pools all treatment arms. Rows below are from specifications that estimate separate effects for each treatment arm. Columns 6 to 10 include controls for personal characteristics (age, gender, type of contract), school characteristics (number of classrooms, number of buildings, land area, number of students, bathroom characteristics, distance to UGEL) and municipality characteristics (altitude, access to electricity, access to drinking water network, access to internet, availability of a bank branch), including dummies for missing observations. Table S1 in the Online Appendix contains the outcome, treatment and control variables' full definitions. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% respectively.

Table A5. Heterogeneous Effects - Expense Report in the Benchmark Experiment

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
SMS	3.855***	4.258***	4.815***	4.919***	4.358***	2.732***	3.286***	2.980***	4.235***	4.451***	4.947***	4.731***	4.429***	3.257***	4.082***	3.909***
	(0.710)	(0.806)	(1.098)	(0.991)	(0.971)	(0.835)	(1.176)	(0.974)	(0.668)	(0.754)	(0.983)	(0.934)	(0.902)	(0.783)	(1.150)	(0.969)
School size X SMS		-0.003								-0.002						
		(0.003)								(0.002)						
School size (P50) X SMS			-1.704								-1.275					
			(1.286)								(1.205)					
Assigned Budget X SMS				-0.000*								-0.000				
				(0.000)								(0.000)				
Assigned Budget (P50) X SMS					-1.254								-0.486			
					(1.367)								(1.297)			
Assigned Budget per Student X SMS						0.007**								0.006**		
						(0.003)								(0.003)		
Assigned Budget per Student (P50) X SMS							0.821								0.216	
							(1.310)								(1.309)	
Rurality X SMS								1.427								0.530
								(1.232)								(1.239)
Controls	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	24,257	24,257	24,257	24,257	24,257	24,257	24,257	24,257	24,257	24,257	24,257	24,257	24,257	24,257	24,257	24,257

Note: Treatment effects and means are reported in percentage points. "SMS" pools all treatment arms. Rows below are from specifications that estimate interactions between treatment and pre-treatment covariates. Columns 9 to 16 include controls for personal characteristics (age, gender, type of contract), school characteristics (number of classrooms, number of buildings, land area, number of students, bathroom characteristics, distance to UGEL) and municipality characteristics (altitude, access to electricity, access to drinking water network, access to internet, availability of a bank branch), including dummies for missing observations. Table S1 in the Online Appendix contains the outcome, treatment and control variables' full definitions. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% respectively.

Table A6. Heterogeneous Effect - Withdrew 99% in the Benchmark Experiment

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
SMS	1.458**	0.225	1.368*	0.675	1.167*	0.347	1.443*	1.137	1.565***	0.956*	1.372**	0.546	1.133**	0.564	1.798**	1.509
	(0.572)	(0.582)	(0.711)	(0.749)	(0.620)	(0.711)	(0.857)	(1.038)	(0.519)	(0.572)	(0.637)	(0.703)	(0.553)	(0.679)	(0.800)	(1.011)
School size X SMS		0.010***								0.005						
		(0.003)								(0.003)						
School size (P50) X SMS			0.170								0.385					
			(1.018)								(0.987)					
Assigned Budget X SMS				0.000								0.000				
				(0.000)								(0.000)				
Assigned Budget (P50) X SMS					0.660								1.095			
					(1.083)								(1.063)			
Assigned Budget per Student X SMS						0.006***								0.006**		
						(0.002)								(0.002)		
Assigned Budget per Student (P50) X SMS							0.018								-0.457	
							(1.036)								(1.014)	
Rurality X SMS								0.501								0.091
								(1.248)								(1.227)
Controls	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,012	21,012	21,012	21,012	21,012	21,012	21,012	21,012	21,012	21,012	21,012	21,012	21,012	21,012	21,012	21,012

Note: Treatment effects and means are reported in percentage points. "SMS" pools all treatment arms. Rows below are from specifications that estimate interactions between treatment and pre-treatment covariates. Columns 9 to 16 include controls for personal characteristics (age, gender, type of contract), school characteristics (number of classrooms, number of buildings, land area, number of students, bathroom characteristics, distance to UGEL) and municipality characteristics (altitude, access to electricity, access to drinking water network, access to internet, availability of a bank branch), including dummies for missing observations. Table S1 in the Online Appendix contains the outcome, treatment and control variables' full definitions. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% respectively.

Table A7. Effects of SMS Campaign on Furniture Quality, Follow-Up Experiment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Variables	**Teaching Desk**	**Teaching Chair**	**Student Chair in Bad Condition**	**Student Desk**	**Board**	**Cupboard**	**Teaching Desk**	**Teaching Chair**	**Student Chair in Bad Condition**	**Student Desk**	**Board**	**Cupboard**
SMS	-0.777	-0.518	0.049	-0.390	0.763	-0.213	-0.894	-0.449	0.035	-0.402	0.737	-0.108
	(0.778)	(0.608)	(0.471)	(0.567)	(0.868)	(1.273)	(0.783)	(0.616)	(0.468)	(0.570)	(0.867)	(1.277)
Descriptive Social Norm	-0.926	-0.949	-0.218	-0.583	0.769	-1.044	-1.087	-0.909	-0.240	-0.616	0.736	-1.032
	(0.874)	(0.657)	(0.542)	(0.636)	(0.975)	(1.375)	(0.887)	(0.664)	(0.543)	(0.647)	(0.981)	(1.378)
Injunctive Social Norm	-0.237	0.422	0.340	-0.287	0.841	0.335	-0.339	0.563	0.322	-0.276	0.833	0.595
	(1.007)	(0.923)	(0.629)	(0.622)	(1.057)	(1.384)	(1.005)	(0.925)	(0.632)	(0.630)	(1.034)	(1.405)
Social Benefit	-0.937	-0.583	0.211	-0.201	0.704	0.514	-1.006	-0.525	0.209	-0.202	0.676	0.655
	(0.796)	(0.676)	(0.566)	(0.704)	(0.967)	(1.516)	(0.806)	(0.686)	(0.559)	(0.692)	(0.965)	(1.512)
Control Mean	4.77	3.74	2.84	3.57	8.21	7.40	4.77	3.74	2.84	3.57	8.21	7.40
Contorls	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,058	6,132	7,285	7,214	7,312	3,777	6,050	6,124	7,277	7,206	7,304	3,771

Note: Treatment effects and means are reported in percentage points. "SMS" pools all treatment arms. Rows below are from specifications that estimate separate effects for each treatment arm. Columns 7 to 12 include controls for personal characteristics (age, gender, type of contract), school characteristics (number of classrooms, number of buildings, land area, number of students, bathroom characteristics, distance to UGEL) and municipality characteristics (altitude, access to electricity, access to drinking water network, access to internet, availability of a bank branch), including dummies for missing observations. Table S1 in the Online Appendix contains the outcome, treatment and control variables' full definitions. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% respectively.

Table A8. Effects of SMS Campaign on Infrastructure Stock, Follow-Up Experiment

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	**Have Hygienic Service**	**Toilets Number**	**Bathroom Sinks Number**	**Have Hygienic Service**	**Toilets Number**	**Bathroom Sinks Number**
SMS	-0.098	0.375	0.016**	-0.076	0.233	0.017**
	(0.300)	(0.478)	(0.008)	(0.297)	(0.377)	(0.008)
Descriptive Social Norm	-0.229	0.229	0.023**	-0.204	0.045	0.024**
	(0.299)	(0.494)	(0.011)	(0.293)	(0.436)	(0.011)
Injunctive Social Norm	-0.051	-0.084	0.008	-0.056	-0.187	0.009
	(0.416)	(0.319)	(0.012)	(0.413)	(0.333)	(0.012)
Social Benefit	0.047	0.873	0.011	0.080	0.762	0.012
	(0.346)	(1.031)	(0.011)	(0.342)	(0.888)	(0.011)
Control mean	99.19	7.43	0.03	99.19	7.43	0.03
Controls	No	No	No	Yes	Yes	Yes
Observations	7,485	7,485	7,485	7,485	7,485	7,485

Note: Treatment effects and means are reported in percentage points. "SMS" pools all treatment arms. Rows below are from specifications that estimate separate effects for each treatment arm. Columns 4 to 6 include controls for personal characteristics (age, gender, type of contract), school characteristics (number of classrooms, number of buildings, land area, number of students, bathroom characteristics, distance to UGEL) and municipality characteristics (altitude, access to electricity, access to drinking water network, access to internet, availability of a bank branch), including dummies for missing observations. Table S1 in the Online Appendix contains the outcome, treatment and control variables' full definitions. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% respectively.

Online Appendix to Accompany "Motivating bureaucrats with non-monetary incentives when state capacity is weak: Evidence from large-scale field experiments in Peru"

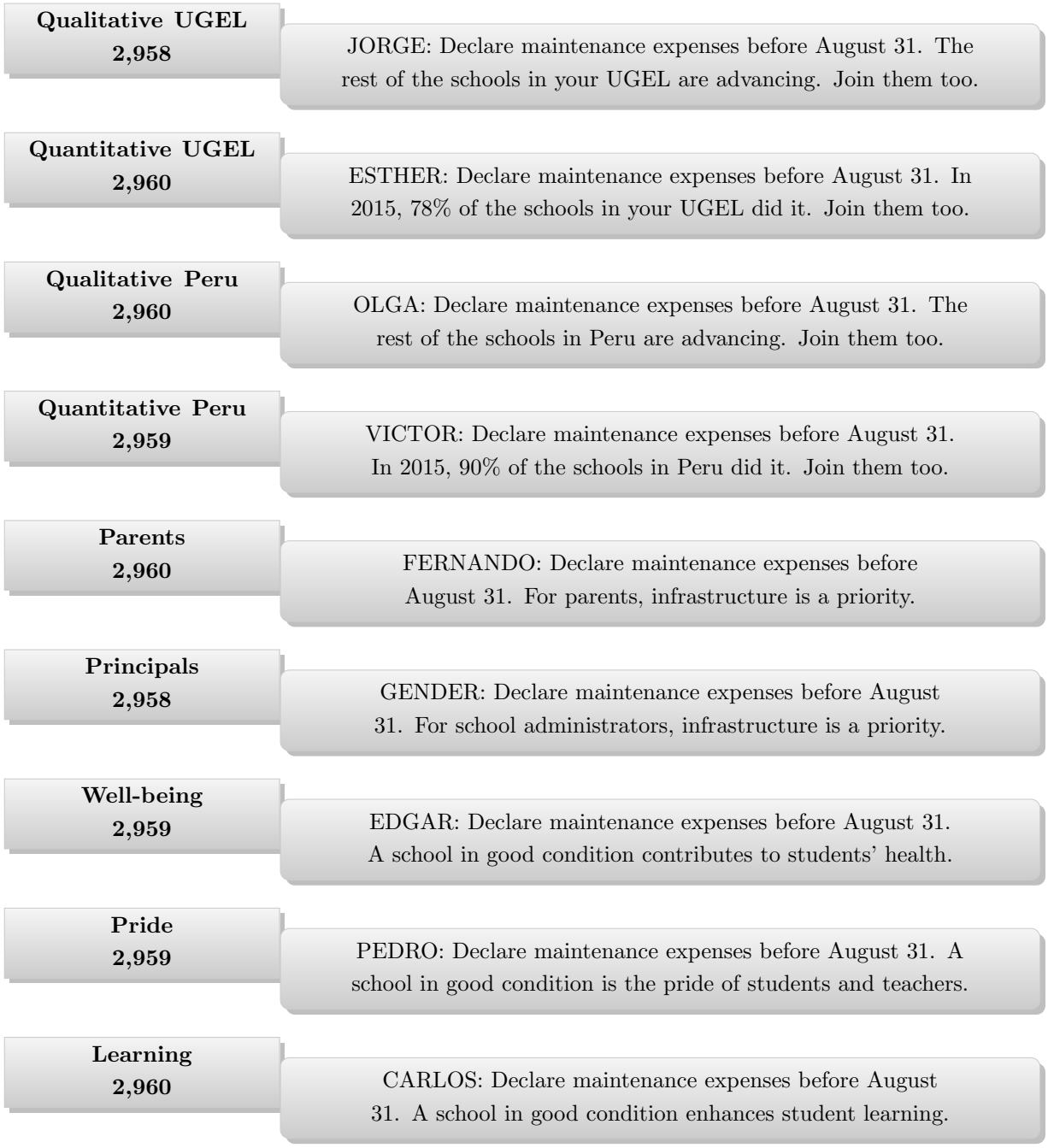
Andrew Dustan
Vanderbilt University

Stanislao Maldonado
Universidad del Rosario

Juan Manuel Hernandez-Agramonte
Innovations for Poverty Action

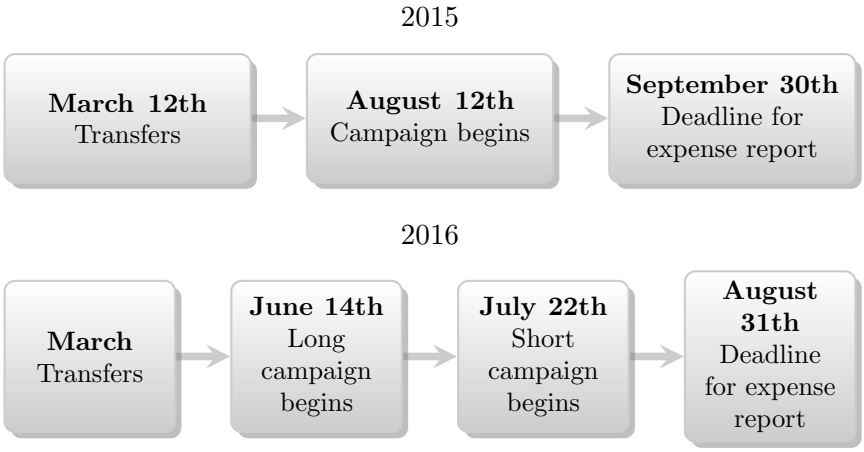
This version: December 22, 2018
For the latest version of this paper, click here.

Figure S1: SMS Content in Follow-Up Experiment in 2016



Note: Number is the sample size of civil servants assigned to the corresponding treatment. Maintenance activity portion of message corresponds to the point in the cycle when the message was sent, as in the Benchmark Experiment (general activities, funds withdrawal, and expense report filing).

Figure S2: Comparison of Intervention Cycles for Benchmark and Follow-Up Experiments

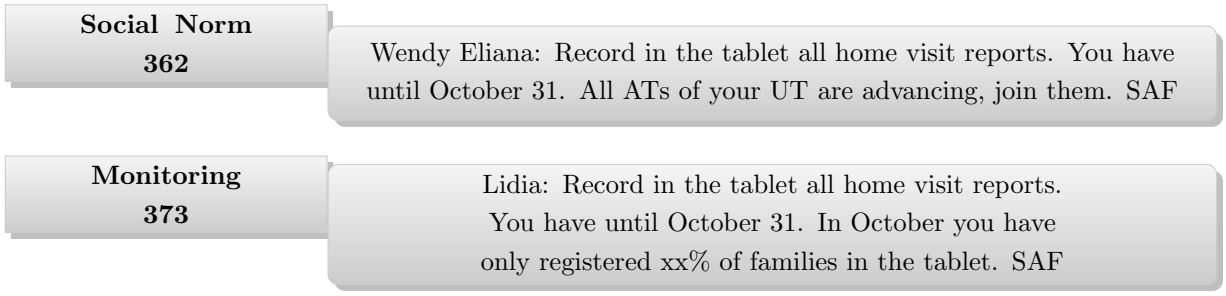


Note: Authors' elaboration. Each square represents a relevant date in the intervention cycle.

Figure S3: Timing of the External Validity Experiment. CUNA MAS Intervention

Note: Authors' elaboration. Each square represents a relevant date in the intervention cycle.

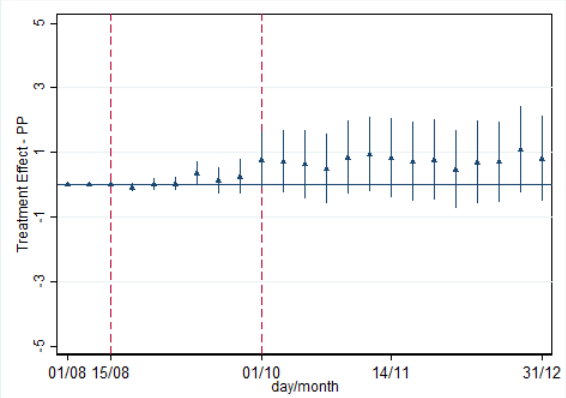
Figure S4: SMS Content in External Validity Experiment. Monthly Messages CUNA MAS



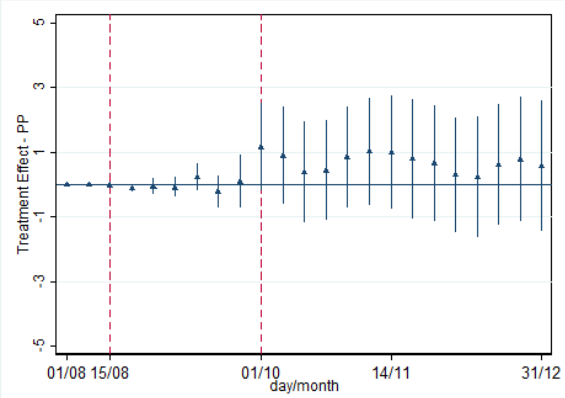
Note: Authors' elaboration. Each message includes the person's name and the deadline to comply with the activity. The rest of the content varies according to the behavioral principle to be emphasized. This example corresponds to the 2nd month of the campaign. All of the messages delivered are described in the Online Appendix (Table S4). Number is the sample size of civil servants (ATs) assigned to the corresponding treatment.

Figure S5: Oversight Report-Benchmark Experiment

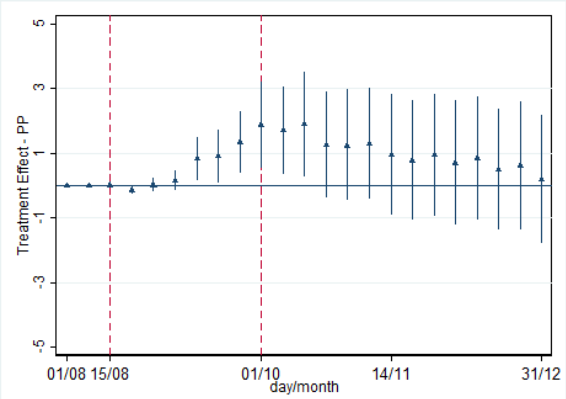
(a) SMS



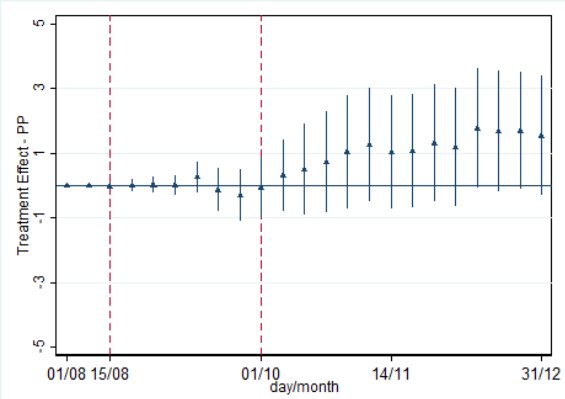
(b) Reminder/Warning



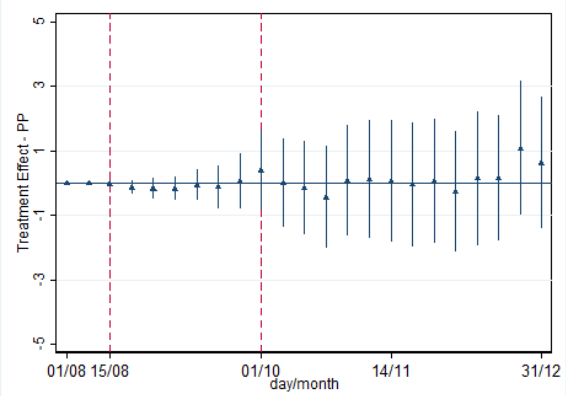
(c) Monitoring



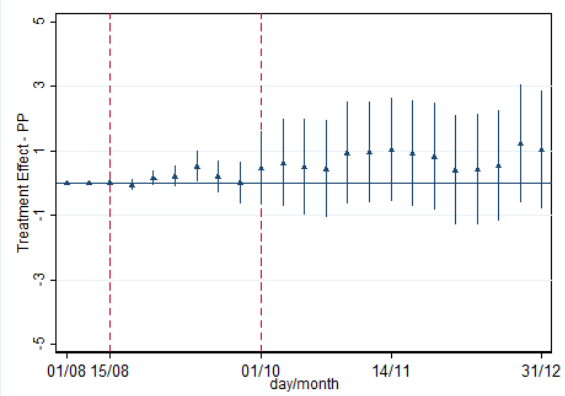
(d) Social Norm



(e) Shaming



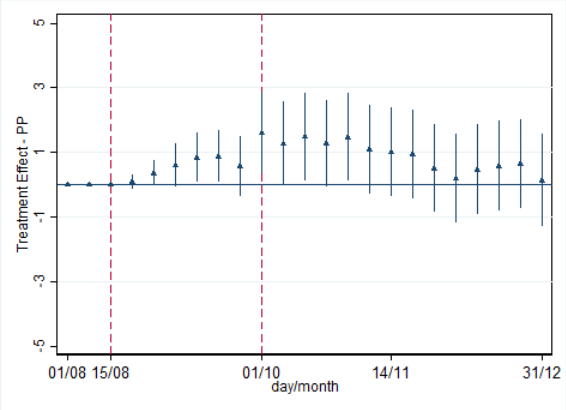
(f) Auditing Threat



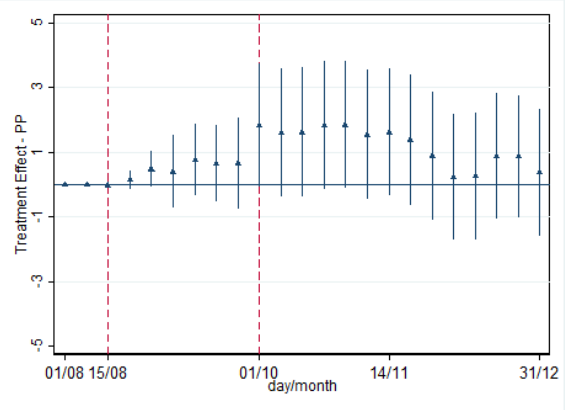
Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Figure S6: Approved Expense Report-Benchmark Experiment

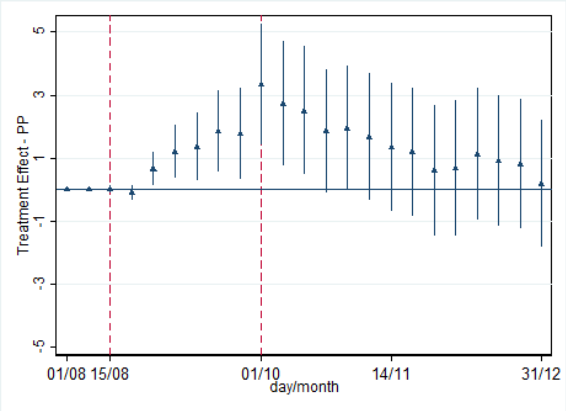
(a) SMS



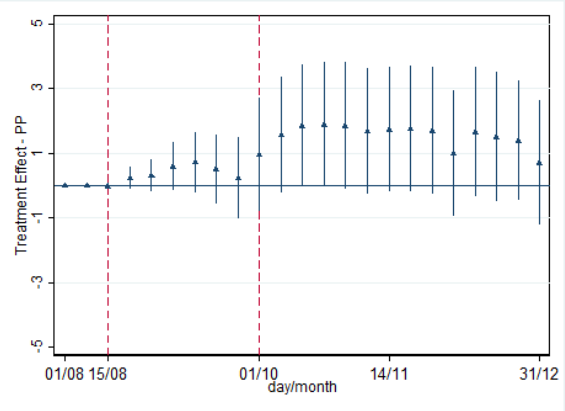
(b) Reminder/Warning



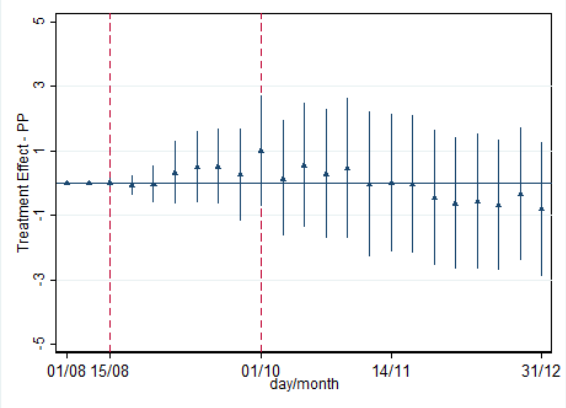
(c) Monitoring



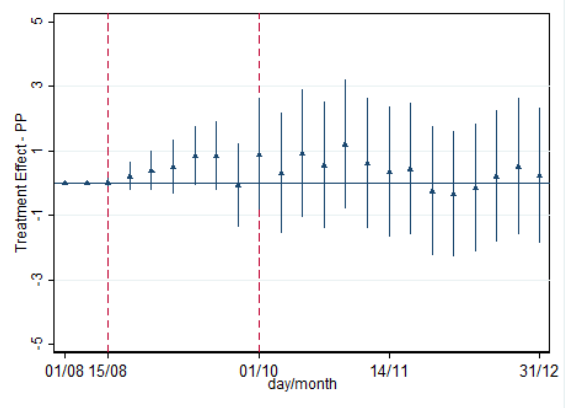
(d) Social Norm



(e) Shaming



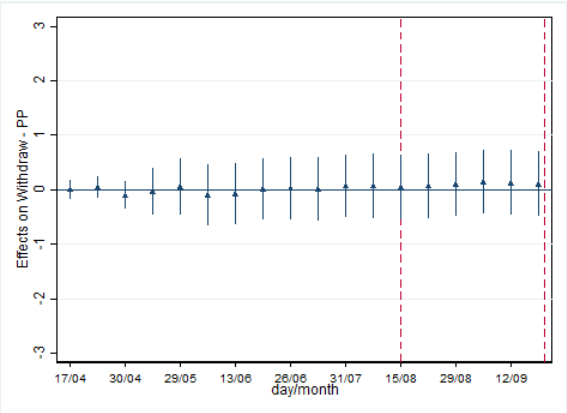
(f) Auditing Threat



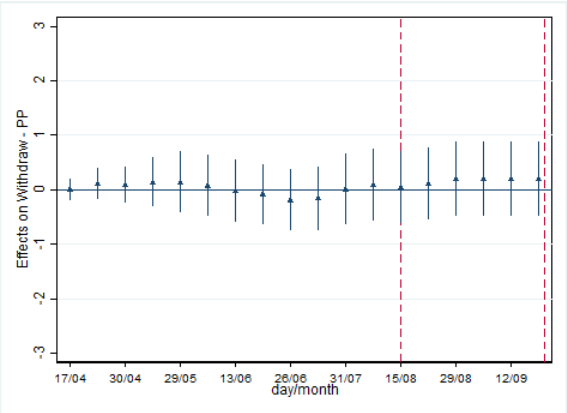
Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Figure S7: Withdrew Something -Benchmark Experiment

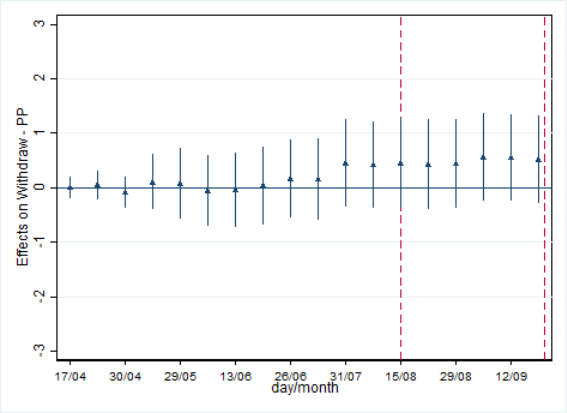
(a) SMS



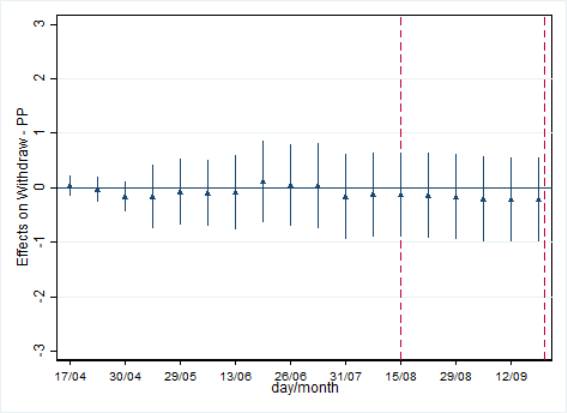
(b) Reminder/Warning



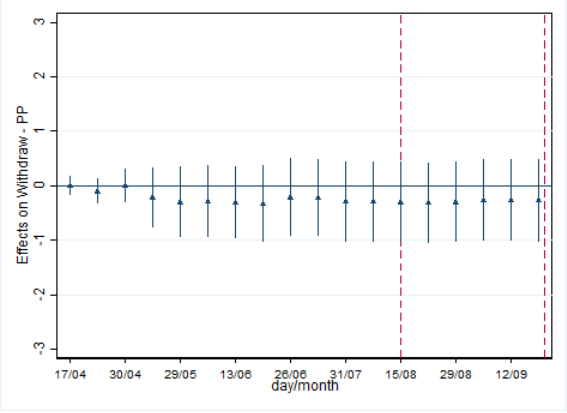
(c) Monitoring



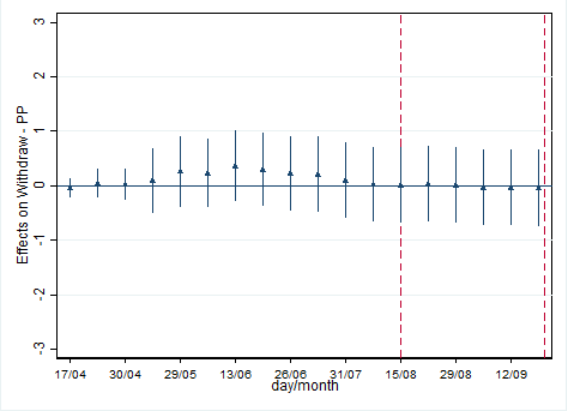
(d) Social Norm



(e) Shaming



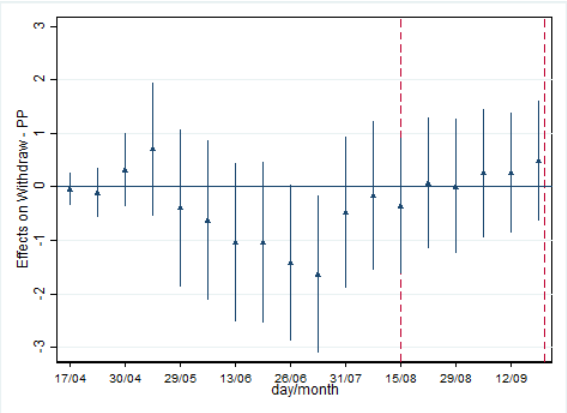
(f) Auditing Threat



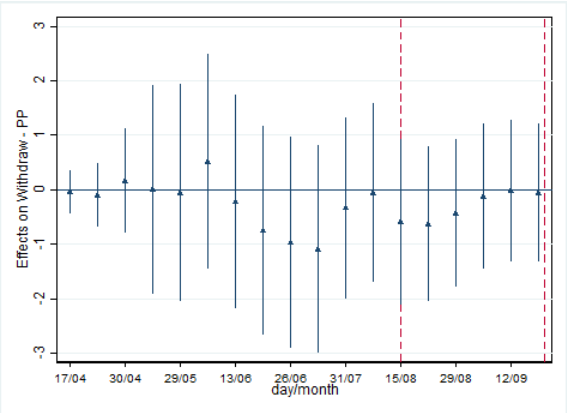
Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Figure S8: Withdrew 50% -Benchmark Experiment

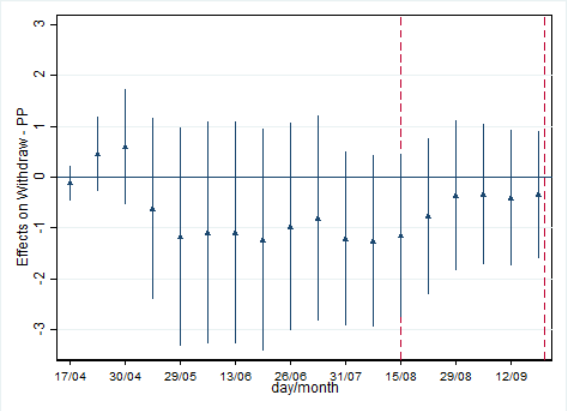
(a) SMS



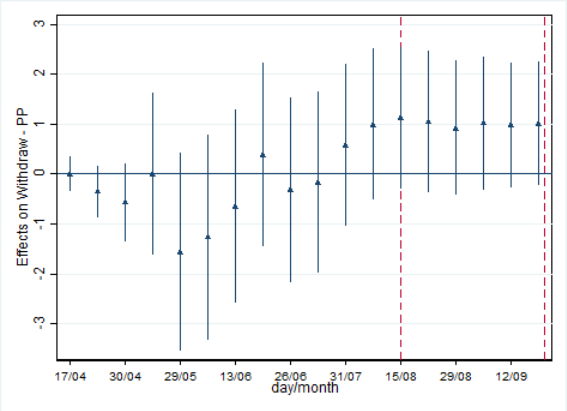
(b) Reminder/Warning



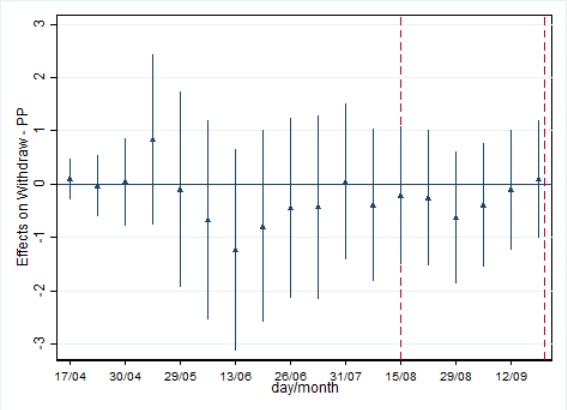
(c) Monitoring



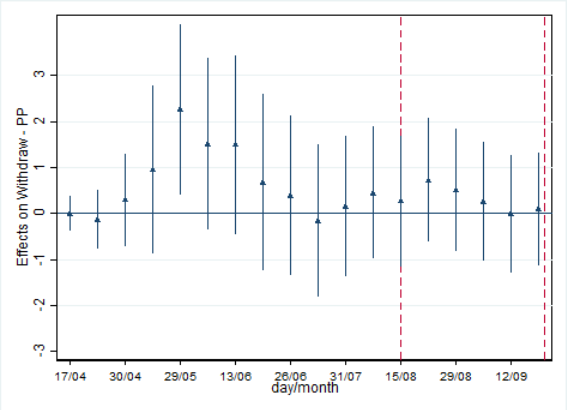
(d) Social Norm



(e) Shaming



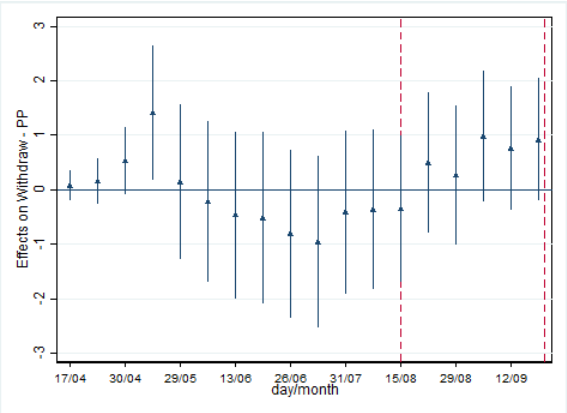
(f) Auditing Threat



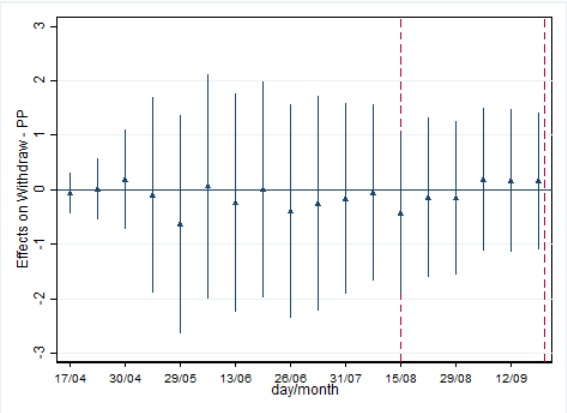
Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Figure S9: Withdrew 80% -Benchmark Experiment

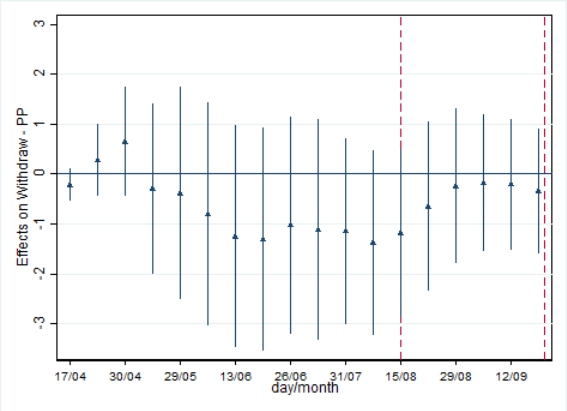
(a) SMS



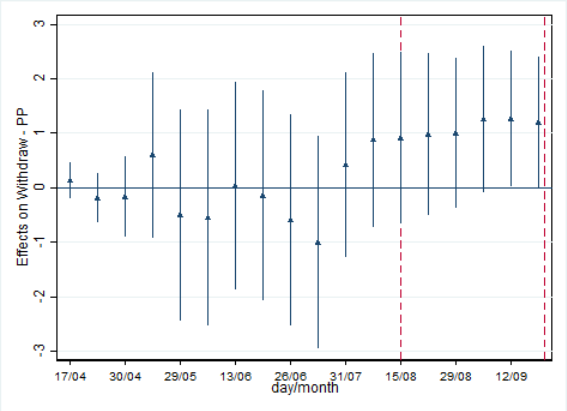
(b) Reminder/Warning



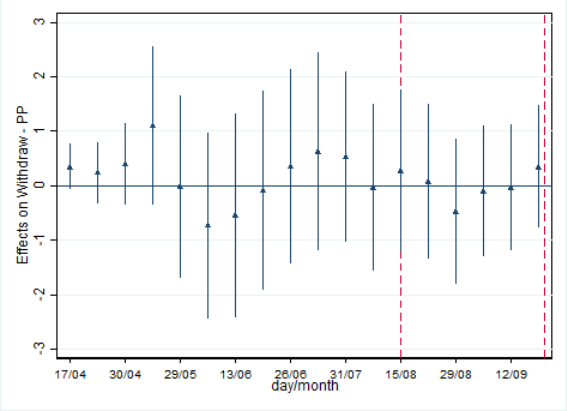
(c) Monitoring



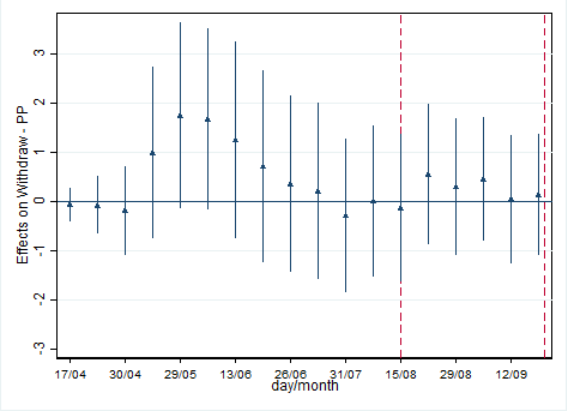
(d) Social Norm



(e) Shaming



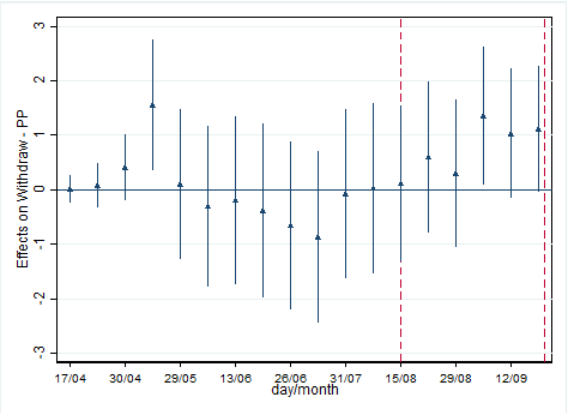
(f) Auditing Threat



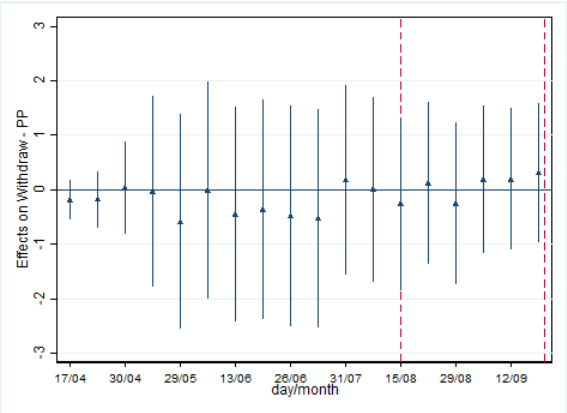
Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Figure S10: Withdrew 95% -Benchmark Experiment

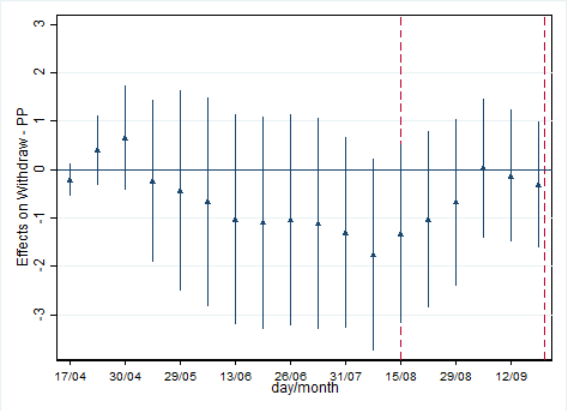
(a) SMS



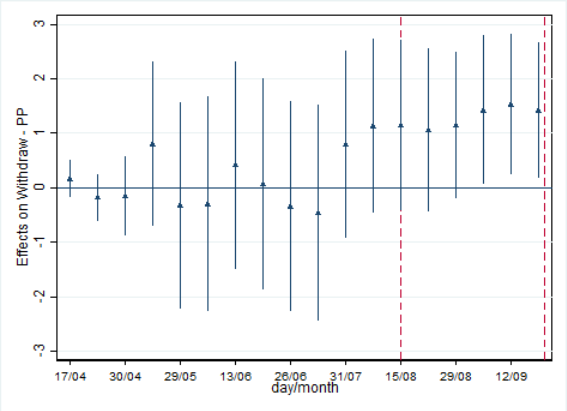
(b) Reminder/Warning



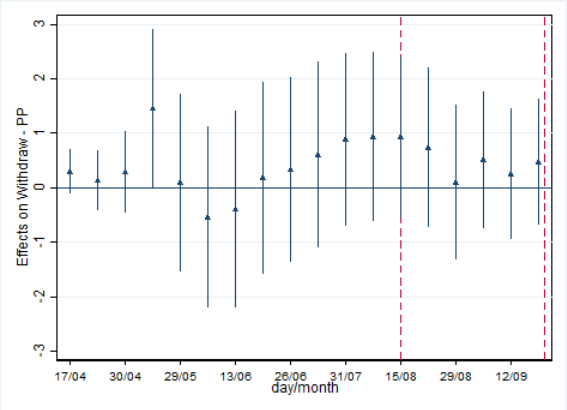
(c) Monitoring



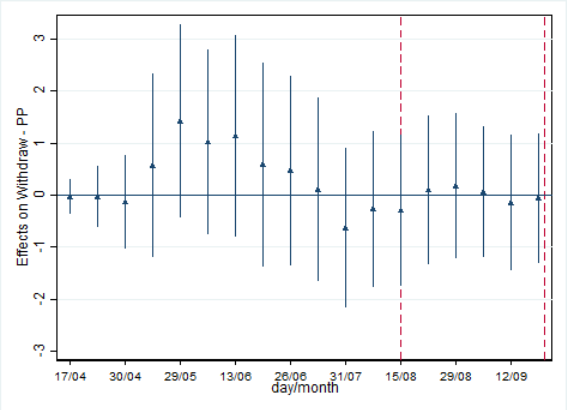
(d) Social Norm



(e) Shaming



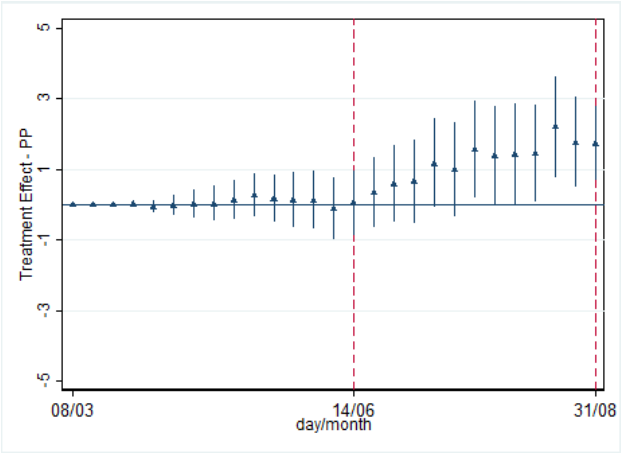
(f) Auditing Threat



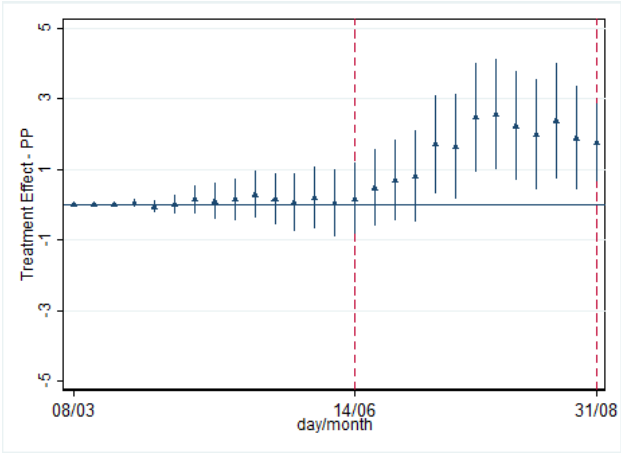
Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Figure S11: Expense Report-Follow-Up Experiment

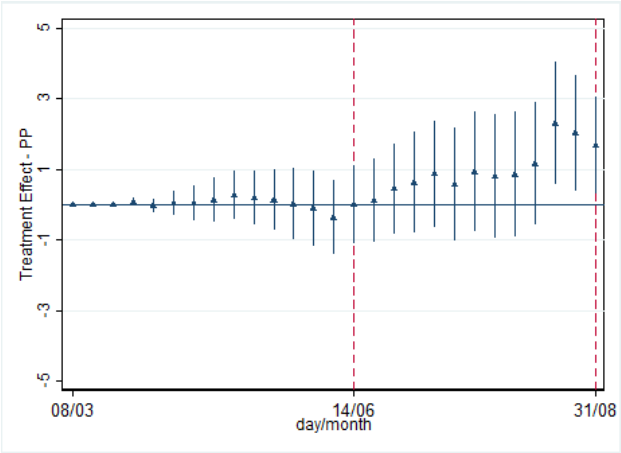
(a) SMS



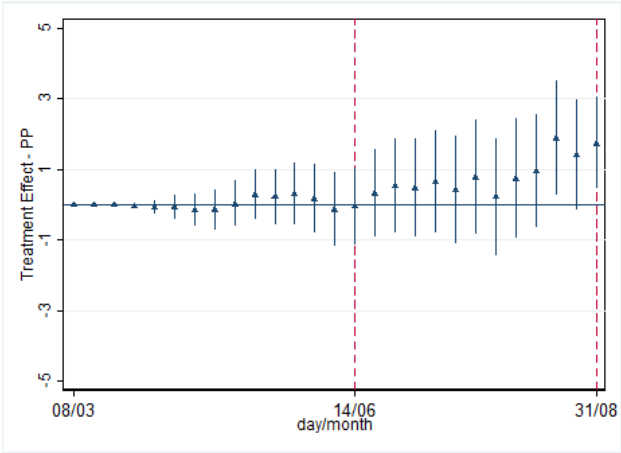
(b) Descriptive Social Norm



(c) Injunctive Social Norm



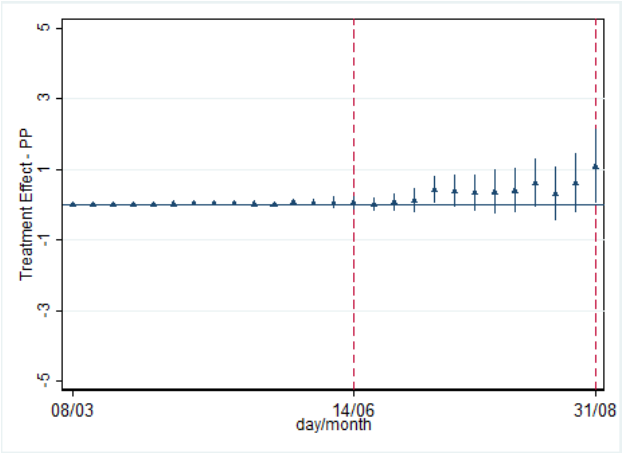
(d) Social Benefit



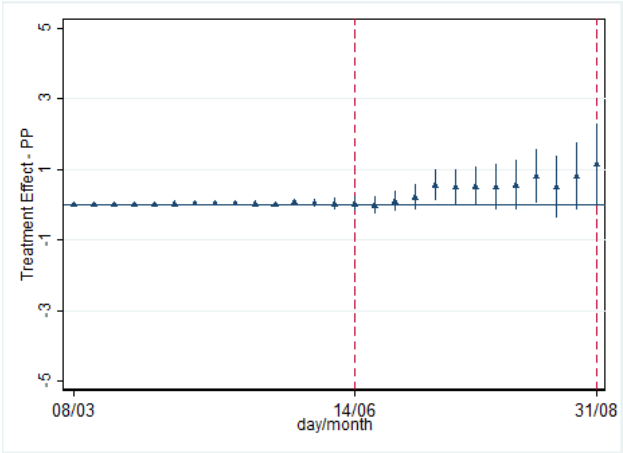
Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Figure S12: Oversight Report-Follow-Up Experiment

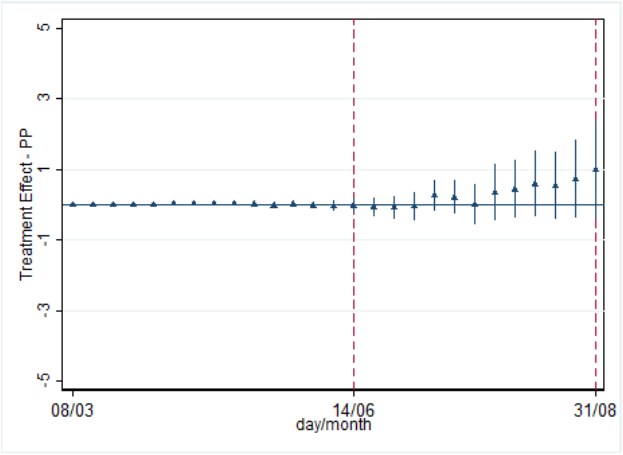
(a) SMS



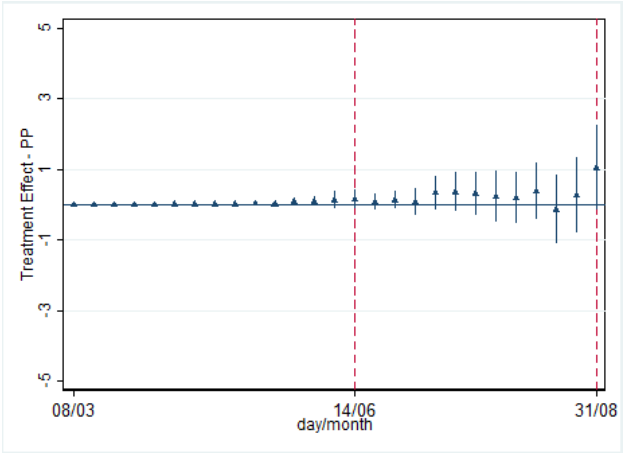
(b) Descriptive Social Norm



(c) Injunctive Social Norm



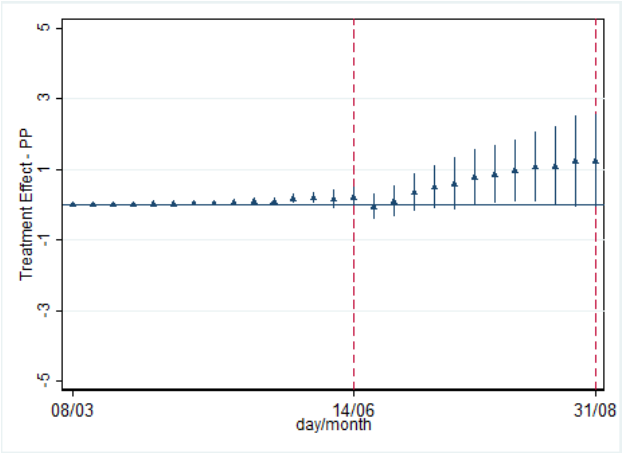
(d) Social Benefit



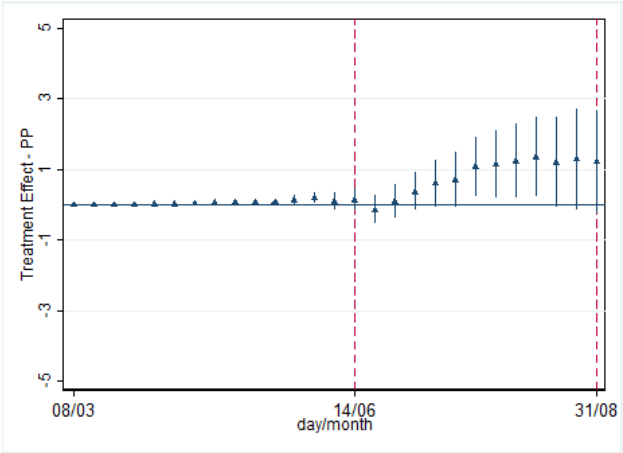
Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Figure S13: Approved Expense Report-Follow-Up Experiment

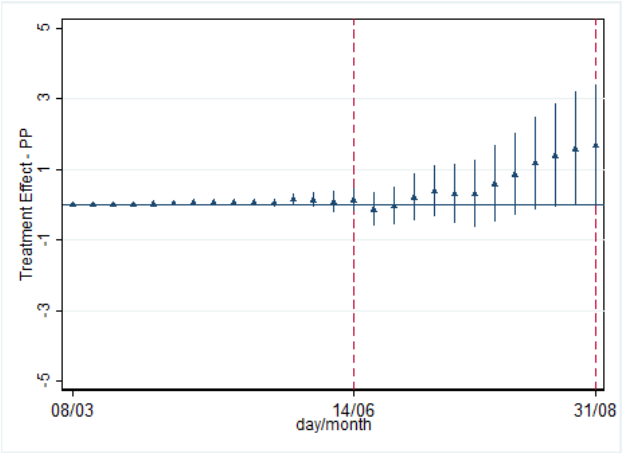
(a) SMS



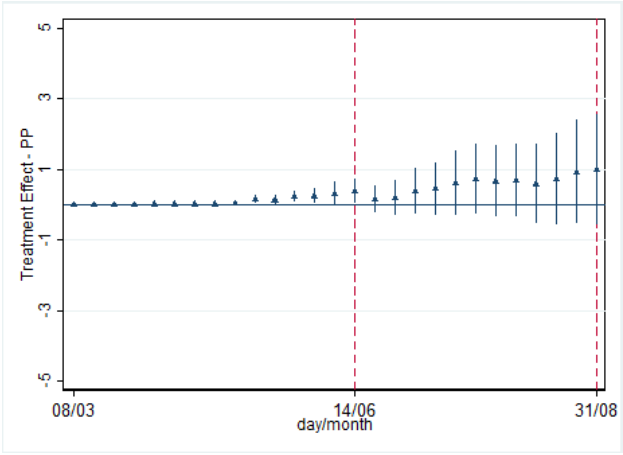
(b) Descriptive Social Norm



(c) Injunctive Social Norm



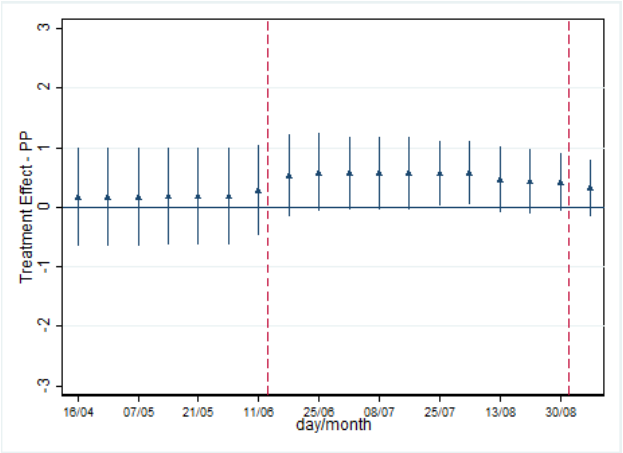
(d) Social Benefit



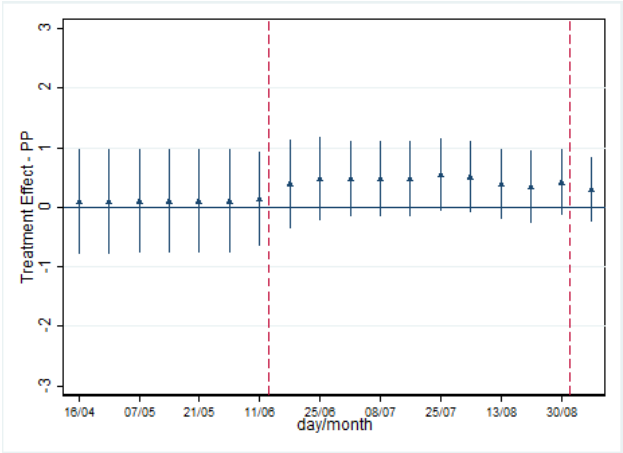
Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Figure S14: Withdrew Something -Follow-Up Experiment

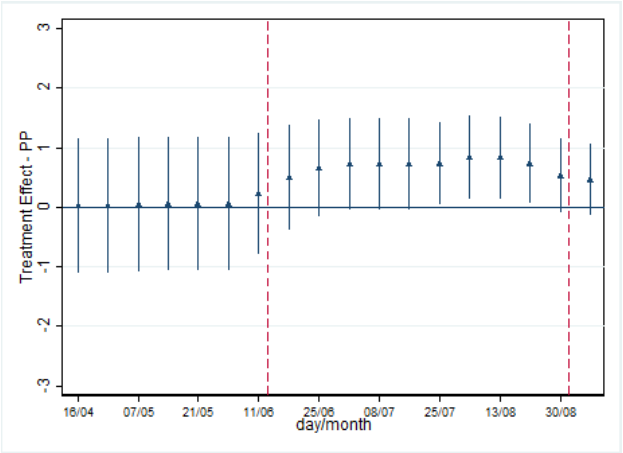
(a) SMS



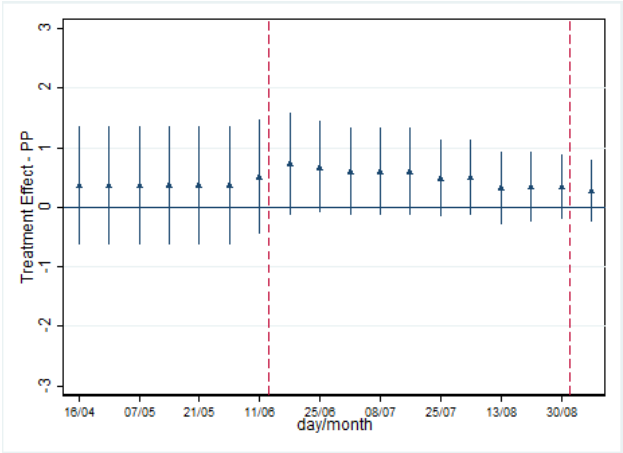
(b) Descriptive Social Norm



(c) Injunctive Social Norm



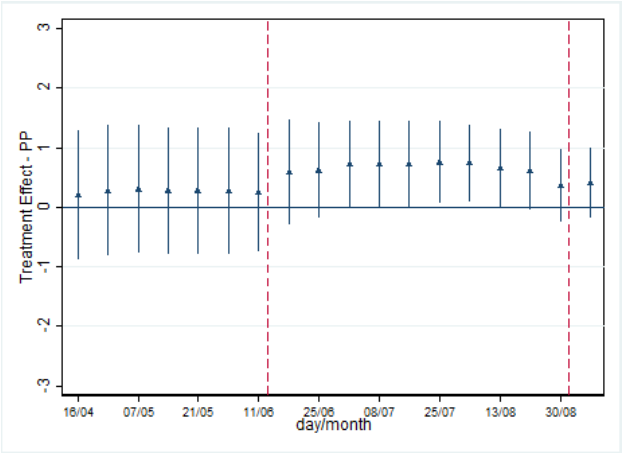
(d) Social Benefit



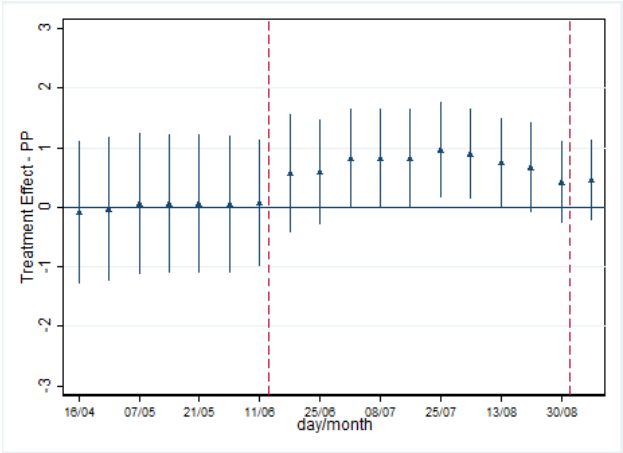
Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Figure S15: Withdrew 50% -Follow-Up Experiment

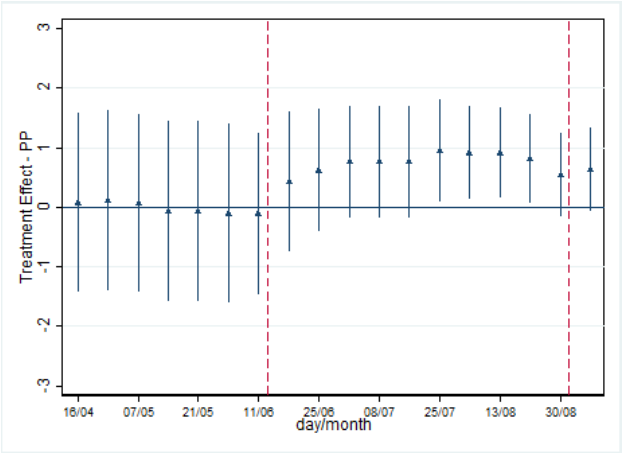
(a) SMS



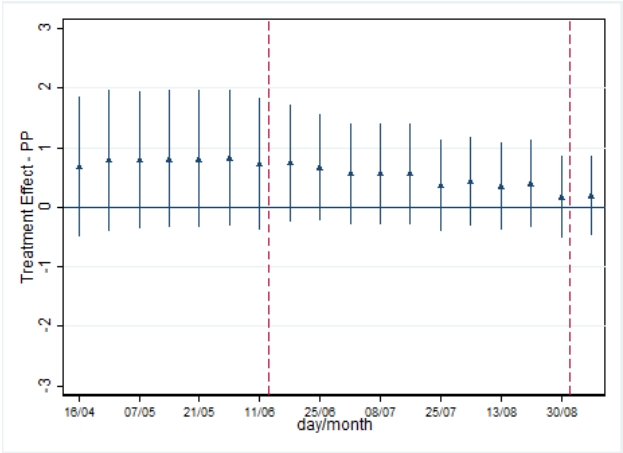
(b) Descriptive Social Norm



(c) Injunctive Social Norm



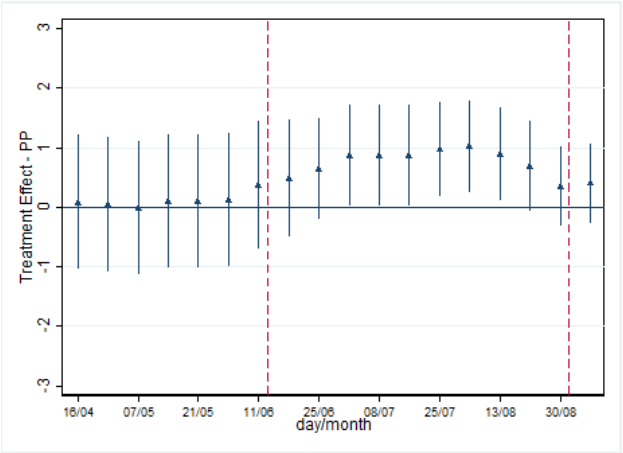
(d) Social Benefit



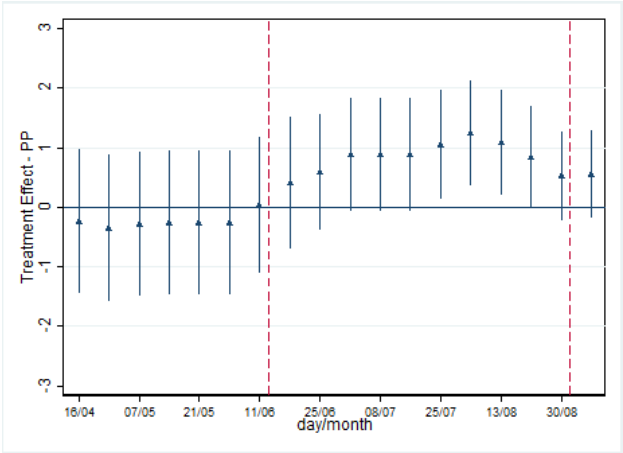
Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Figure S16: Withdrew 80% -Follow-Up Experiment

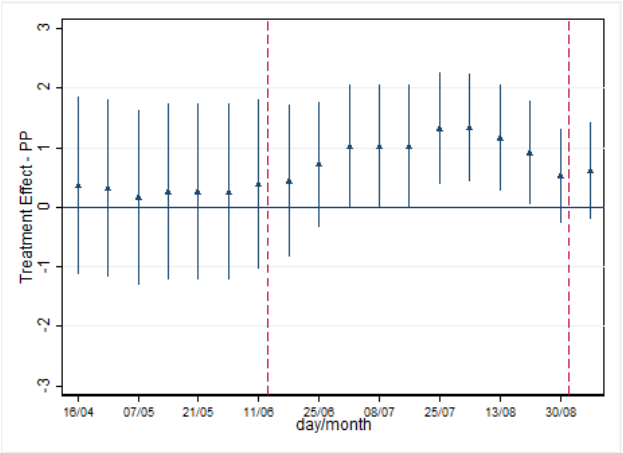
(a) SMS



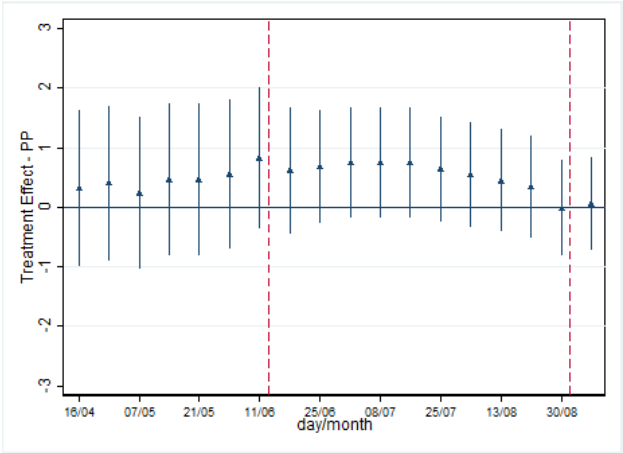
(b) Descriptive Social Norm



(c) Injunctive Social Norm



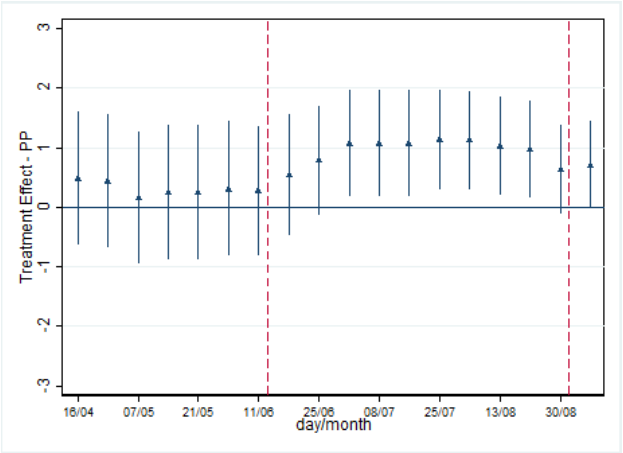
(d) Social Benefit



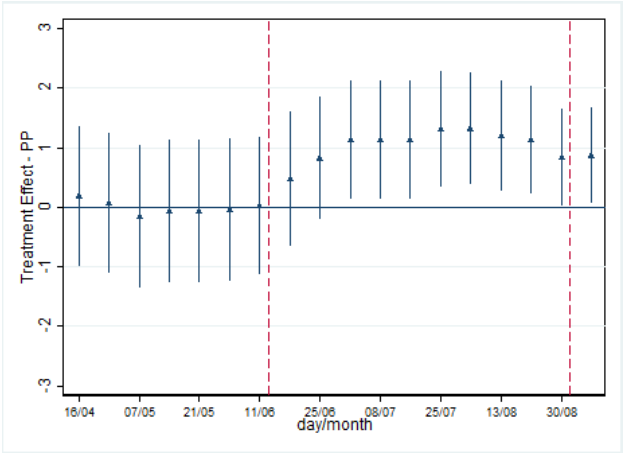
Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Figure S17: Withdrew 95% -Follow-Up Experiment

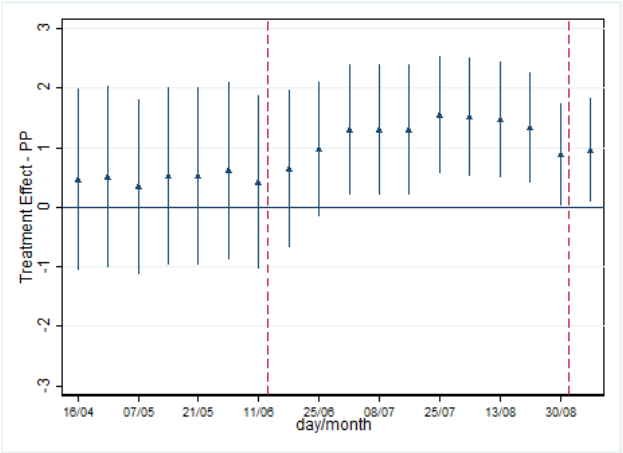
(a) SMS



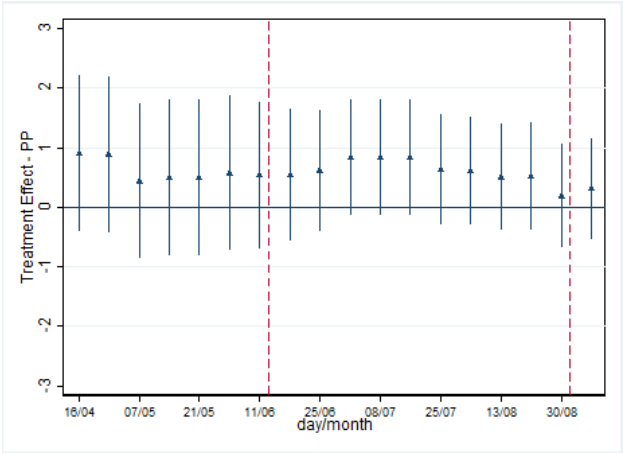
(b) Descriptive Social Norm



(c) Injunctive Social Norm



(d) Social Benefit



Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Equation 1 with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Table S1. Full List of Variable Definitions

Variable	Definition
Panel A: Pre-treatment Outcomes	
Submitted to Maintenance Committee at Week 20	Percentage of schools that had formed a maintenance comittee at week 20 after the maintance funds had been assigned.
Submitted to Oversight Committee at Week 20	Percentage of schools that had formed an oversight committee at week 20 after the maintenance funds had been assigned.
Submitted Technical Form at Week 20	Percentage of schools that had submitted a technical form to their corresponding UGEL at week 20 after the maintenance funds had been assigned.
Submitted Commitment Act at Week 20	Percentage of schools that had submitted a commitment act at week 20 after the maintenance funds had been assigned.
Submitted Expense Report at Week 20	Percentage of schools that had submitted an expense report at week 20 after the maintenance funds had been assigned.
Submitted Oversight Report at Week 20	Percentage of schools that had submitted an oversight report at week 20 after the maintenance funds had been assigned.
With Approved Expense Report at Week 20	Percentage of schools that had their expenses report approved by their corresponding UGEL at week 20 after the maintenance funds had been assigned.
Bank Balance at 26/06/2016	Bank Balance at 26/06/2016.
Panel B: Outcomes	
Submitted Commitment Act at Week 30	Percentage of schools that had submitted a commitment act at week 30 after the maintenance funds had been assigned.
Submitted Expense Report at Week 30	Percentage of schools that had submitted an expense report at week 30 after the maintenance funds had been assigned.
Submitted Oversight Report at Week 30	Percentage of schools that had submitted an oversight report at week 30 after the maintenance funds had been assigned.
With Approved Expense Report at Week 30	Percentage of schools that had their expense report approved by their corresponding UGEL at week 30 after the maintenance funds had been assigned.
Withdrew Something	Percentage of maintenance civil servants who withdrew any positive amount from the assigned funds.
Withdrew 50%	Percentage of maintenance civil servants who withdrew at least 50% from the assigned funds.
Withdrew 80%	Percentage of maintenance civil servants who withdrew at least 80% from the assigned funds.
Withdrew 95%	Percentage of maintenance civil servants who withdrew at least 95% from the assigned funds.
Withdrew 99%	Percentage of maintenance civil servants who withdrew at least 99% from the assigned funds.

Table S1. Full List of Variable Definitions (Continued)

Variable	Definition
Panel C: Maintenance CS Characteristics	
Sex	Dummy that takes the value of 1 if the chosen maintenance civil servant is male and 0 otherwise.
Age	Age of the chosen maintenance civil servant.
Appointed Maintenance CS	Dummy that takes the value of 1 if the maintenance civil servant was appointed and 0 otherwise.
Hired Maintenance CS	Dummy that takes the value of 1 if the maintenance civil servant was hired and 0 otherwise.
Allocation Transfer	Total amount of money that was allocated to the maintenance civil servant.
Panel D: School Characteristics	
Classrooms	Total number of classrooms in the school.
Students	Total number of students in the school.
Bathroom Connected to Public Drainage System	Dummy that takes the value of 1 if the school's bathrooms were connected to the public drainage system and 0 otherwise.
Bathroom Connected to Septic Tank	Dummy that takes the value of 1 if the school's bathrooms were connected to a septic tank and 0 otherwise.
Bathroom Connected to a Black Well	Dummy that takes the value of 1 if the school's bathrooms were connected to a black well and 0 otherwise.
Bathroom Connected to River, Ditch or Canal	Dummy that takes the value of 1 if the school's bathrooms were connected to a river, ditch or canal and 0 otherwise.
No Bathroom	Dummy that takes the value of 1 if the school didnt have bathrooms and 0 otherwise.
Total Land Area	Total school area in squared meters.
Fully fenced	Dummy that takes the value of 1 if the school was fully fenced and 0 otherwise.
Partially Fenced	Dummy that takes the value of 1 if the school had a partial fence and 0 otherwise.
Unfenced	Dummy that takes the value of 1 if the school was unfenced and 0 otherwise.
Number Educ-Admin Spaces	Total number of educative/administrative spaces (classrooms, computer rooms, laboratories, workshops, libraries, teacher's rooms and gyms).
Number of Buildings	Total number of independent buildings or pavilions in the school where an independent building or pavilion is defined to be an edification with one or more classrooms with common walls and/or roofs distributed among one or more floors.
Average Leaks in Pavilions	Average number of leaks, fissures and cracks in the classrooms.
Average Leaks	Average number of leaks in the classrooms.

Table S1. Full List of Variable Definitions (Continued)

Variable	Definition
Panel E: District Characteristics	
Altitude	Altitude of the district measured in meters above the sea level.
Area	Dummy that takes the value of 1 if the district is located in a rural area and 0 otherwise.
Electricity	Dummy that takes the value of 1 if the district has access to electricity.
Public Drinking Water Network	Dummy that takes the value of 1 if the district has a drinking water network.
Public Drainage Network	Dummy that takes the value of 1 if the district has a drainage network.
Internet Cafe	Dummy that takes the value of 1 if the district has access to internet.
Bank Branch	Dummy that takes the value of 1 if there is a banking institution in the district.
Panel F: Treatments	
SMS	Dummy that takes the value of 1 if the maintenance civil servant received any kind of SMS message.
Reminder/Warning	Dummy that takes the value of 1 if the maintenance civil servant received an SMS message of the Reminder/Warning type.
Social Norm	Dummy that takes the value of 1 if the maintenance civil servant received an SMS message of the Social Norm type.
Monitoring	Dummy that takes the value of 1 if the maintenance civil servant received an SMS message of the Monitoring type.
Shaming	Dummy that takes the value of 1 if the maintenance civil servant received an SMS message of the Shaming type.
Auditing Threat	Dummy that takes the value of 1 if the maintenance civil servant received an SMS message of the Auditing Threat type.

Table S2. SMS in the Benchmark Experiment

	Treatment	SMS
Maintenance Activities	Reminder/Warning	YRMA: REMEMBER, perform maintenance activities according to the file registered in Wasichay. For more details, visit www.pronied.gob.pe.
	Monitoring	LUCILA: Perform maintenance activities according to the file registered in Wasichay. You have pending activities.
	Social Norm	BENJAMIN: Perform maintenance activities according to the file registered in Wasichay. The rest of schools in your UGEL are advancing. You are behind.
	Shaming	ADRIAN: Perform maintenance activities according to the file registered in Wasichay. We will publish the names of schools and civil servants that do not comply.
	Auditing Threat	KARINA: Perform maintenance activities according to the file registered in Wasichay.We will visit your school to supervise activities.
Withdrawal of Allocation Transfer	Reminder/Warning	YRMA: REMEMBER, withdraw the allocated transfer for maintenance. For more details consult the specialist of your UGEL.
	Monitoring	LUCILA: Withdraw the allocated transfer for maintenance.
	Social Norm	BENJAMIN: Withdraw the allocated transfer for maintenance. 89% of schools in your UGEL have already withdrawn the allocated amount. You are behind.
	Shaming	ADRIAN: Withdraw the allocated transfer for maintenance. We will publish the names of schools and civil servants that do not comply.
	Auditing Threat	KARINA: Withdraw the allocated transfer for maintenance. We will visit your school to supervise activities.
Declare Expenditure	Reminder/Warning	YRMA: ALERT! Declare maintenance expenses before September 30th. For more details consult the specialist of your UGEL.
	Monitoring	LUCILA: Declare maintenance expenses before September 30th. You have S/.2000 still undeclared in the Wasichay system.
	Social Norm	BENJAMIN: Declare maintenance expenses before September 30th. The rest of the schools in your UGEL are advancing. You are behind.
	Shaming	ADRIAN: Declare maintenance expenses before September 30th. We will publish the names of schools and civil servants that do not.
	Auditing Threat	KARINA: Declare maintenance expenses before September 30th. We will visit your school to supervise activities.
Declare all the Allocated Transfer	Reminder/Warning	YRMA: URGENT! Declare all the allocated transfer before September 30th. For more details consult the specialist of your UGEL.
	Monitoring	LUCILA: Declare all the allocated transfer before September 30th. You have S/.2000 without declaring in the Wasichay system.
	Social Norm	BENJAMIN: Declare all the allocated transfer before September 30th. The rest of schools in your UGEL are advancing. You are behind.
	Shaming	ADRIAN: Declare all the allocated transfer before September 30th. We will publish the names of schools and civil servants that do not comply.
	Auditing Threat	KARINA: Declare all the allocated transfer before September 30th. We will visit your school to supervise activities.

Note: The compliance percentages and bank balance amounts are examples. Actual messages corresponded to each civil servant's case.

Table S3. SMS in the Follow-Up Experiment

	Treatment	**SMS**
Maintenance Activities	Qualitative UGEL	JORGE: Performs maintenance actions according to the file registered in Wasichay. The rest of the schools of your UGEL are advancing. Add yourself too.
	Quantitative UGEL	ESTHER: Perform maintenance actions according to the file registered in Wasichay. In 2015, 78% of schools in your UGEL did it. Join them too.
	Qualitative Peru	OLGA: Perform maintenance actions according to the file registered in Wasichay. The rest of the schools in Peru are advancing. Join them too.
	Quantitative Peru	VICTOR: Perform maintenance actions according to the file registered in Wasichay. In 2015, 90% of schools in Peru did it. Join them too.
	Parents	FERNANDO: Perform maintenance actions according to the file registered in Wasichay. For parents, infrastructure is a priority.
	Principals	GENDER: Perform maintenance actions according to the file registered in Wasichay. For school administrators, infrastructure is a priority.
	Well-being	EDGAR: Perform maintenance actions according to the file registered in Wasichay. A school in good condition contributes to student health.
	Pride	PEDRO: Perform maintenance actions according to the file registered in Wasichay. A school in good condition is the pride of students and teachers.
	Learning	CARLOS: Perform maintenance actions according to the file registered in Wasichay. A school in good condition enhances student learning.
Withdrawal of Allocated Transfer	Qualitative UGEL	JORGE: Withdraw all the allocated transfer for maintenance of your school. The rest of the schools of your UGEL are advancing. Join them too.
	Quantitative UGEL	ESTHER: Withdraw all the allocated transfer for maintenance of your school. In 2015, 94% of schools in your UGEL did it. Join them too.
	Qualitative Peru	OLGA: Withdraw all the allocated transfer for maintenance of your school. The rest of the schools in Peru are advancing. Join them too.
	Quantitative Peru	VICTOR: Withdraw all the allocated transfer for maintenance of your school. In 2015, 89% of schools in Peru did it. Join them too.
	Parents	FERNANDO: Withdraw all the allocated transfer for maintenance of your school. For parents, infrastructure is a priority.
	Principals	GENDER: Withdraw all the allocated transfer for maintenance of your school. For school administrators, infrastructure is a priority.
	Well-being	EDGAR: Withdraw all the allocated transfer for maintenance of your school. A school in good condition contributes to student health.
	Pride	PEDRO: Withdraw all the allocated transfer for maintenance of your IE. A school in good condition is the pride of students and teachers.
	Learning	CARLOS: Withdraw all the allocated transfer for maintenance of your school. A school in good condition favors student learning.

Table S3. SMS in the Follow-Up Experiment (Continued)

	Treatment	SMS
Declare Expenditure	Qualitative UGEL	JORGE: Declare maintenance expenses before August 31. The rest of the schools in your UGEL are advancing. Join them too.
	Quantitative UGEL	ESTHER: Declare maintenance expenses before August 31. In 2015, 78% of schools in your UGEL did it. Join them too.
	Qualitative Peru	OLGA: Declare maintenance expenses before August 31. The rest of the schools in Peru are advancing. Join them too.
	Quantitative Peru	VICTOR: Declare maintenance expenses before August 31. In 2015, 90% of schools in Peru did it. Join them too.
	Parents	FERNANDO: Declare maintenance expenses before August 31. For parents, infrastructure is a priority.
	Principals	GENDER: Declare maintenance expenses before August 31. For school administrators, infrastructure is a priority.
	Well-being	EDGAR: Declare maintenance expenses before August 31. A school in good condition contributes to student health.
	Pride	PEDRO: Declare maintenance expenses before August 31. A school in good condition is the pride of students and teachers.
	Learning	CARLOS: Declare maintenance expenses before August 31. A school in good condition enhances student learning.
Declare all the Allocated Transfer	Qualitative UGEL	JORGE: Declare all the allocated transfer before August 31. The rest of the schools of your UGEL are advancing. Join them too.
	Quantitative UGEL	ESTHER: Declare all the allocated transfer before August 31. In 2015, 78% of schools in your UGEL did it. Join them too.
	Qualitative Peru	OLGA: Declare all the allocated transfer before August 31. The rest of the schools in Peru are advancing. Join them too.
	Quantitative Peru	VICTOR: Declare all the allocated transfer before August 31. In 2015, 90% of schools in Peru did it. Join them too.
	Parents	FERNANDO: Declare all the allocated transfer before August 31. For parents, infrastructure is a priority.
	Principals	GENDER: Declare all the allocated transfer before August 31. For school administrators, infrastructure is a priority. PRONIED
	Well-being	EDGAR: Declare all the allocated transfer before August 31. A school in good condition contributes to student health.
	Pride	PEDRO: Declare all the allocated transfer before August 31. A school in good condition is the pride of students and teachers.
	Learning	CARLOS: Declare all the allocated transfer before August 31. A school in good condition enhances student learning. PRONIED

Note: The compliance percentages and bank balance amounts are examples. Actual messages corresponded to each civil servant's case.

Table S4. SMS in External Validity (CUNA MAS) Experiment

	Treatment	SMS
Planning Meeting Sep/Oct/Nov/Jun 12th	Social Norm	Wendy Eliana: At the planning meeting, record in the tablet the visit reports to date. All ATs in your UT are advancing, join them. SAF
	Monitoring	Lidia: At the planning meeting, record in the tablet the visit reports to date. In October you only registered xx% of families in the tablet. SAF
Reminder Deadline Sep/Oct/Nov/Jun 21st	Social Norm	Wendy Eliana: Record in the tablet all home visit reports. You have until October 31. All ATs in your UT are advancing, join them. SAF
	Monitoring	Lidia: Record in the tablet all home visit reports. You have until October 31. In October you only registered xx% of families in the tablet. SAF
Reminder Deadline Sep/Oct/Nov/Jun 26th	Social Norm	Wendy Eliana: Record in the tablet all home visit reports. You have until October 31. All ATs in your UT are advancing, join them. SAF
	Monitoring	Lidia: Record in the tablet all home visit reports. You have until October 31. In October you only registered xx% of families in the tablet. SAF

Note: Authors' elaboration.

Table S5. Descriptive Statistics for Follow-Up Experiment 2016

Variables	Mean	Std. Dev.	Min	Max	N
Pre-treatment Outcomes					
Submitted to Maintenance Committee at Week 15	0.883	0.321	0	1	31,947
Submitted to Oversight Committee at Week 15	0.882	0.322	0	1	31,947
Submitted Technical Form at Week 15	0.767	0.422	0	1	31,947
Submitted Commitment Act at week 15	0.742	0.438	0	1	31,947
Submitted Expense Report at Week 15	0.108	0.310	0	1	31,947
Submitted Oversight Report at Week 15	0.006	0.075	0	1	31,947
With Approved Expense Report at Week 15	0.013	0.114	0	1	31,947
Bank Balance at 14/05/2016	2,691	3,929	0	30,010	29,923
Bank Balance at 11/06/2016	1,554	3,182	0	30,015	31,733
Outcomes					
Submitted Commitment Act at Week 43	0.903	0.296	0	1	31,947
Submitted Expense Report at Week 43	0.821	0.384	0	1	31,947
Submitted Oversight Report at Week 43	0.183	0.386	0	1	31,947
With Approved Expenses Report at Week 43	0.402	0.490	0	1	31,947
Withdrew Something	0.9651	0.1836	0	1	31,947
Withdrew 50%	0.9376	0.2419	0	1	31,947
Withdrew 80%	0.9206	0.2703	0	1	31,947
Withdrew 95%	0.9096	0.2867	0	1	31,947
Withdrew 99%	0.9061	0.2917	0	1	31,947
Maintenance CS Characteristics					
Sex (% Men)	0.463	0.499	0	1	30,296
Age	44.0	12.5	0	116.4	31,947
Allocation Transfer	6,981	3,770	0	30,000	31,947
School Characteristics					
Classrooms	4.9	5.5	1	76	31947
Students	85.2	575.1	0	83,032	24,268
Bathroom Connected to Public Drainage System	0.341	0.474	0	1	31,947
Bathroom Connected to Septic Tank	0.241	0.428	0	1	31,947
Bathroom Connected to a Black Well	0.271	0.444	0	1	31,947
Bathroom Connected to River, Ditch or Canal	0.029	0.169	0	1	31,947
No Bathroom	0.068	0.252	0	1	31,947
Total Land Area	6,526	55,689	0	7,381,000	31,947
Fully Fenced	0.295	0.456	0	1	31,947
Partially Fenced	0.242	0.428	0	1	31,947
Unfenced	0.412	0.492	0	1	31,947
Number Educ-Admin Spaces	7.1	8.4	0	191	31,947
Number of Buildings	2.0	2.1	0	49	31,947
Average Leaks in Pavilions	1.1	1.4	0	40	31,947
Average Leaks	1.0	1.3	0	25	31,947
District Characteristics					
Altitude	2,044	1,464	0	5,131	31,947
Area (% Rural)	0.348	0.476	0	1	31,947
Electricity	0.810	0.393	0	1	30,344
Public Drinking Water Network	0.646	0.478	0	1	30,343
Public Drainage Network	0.366	0.482	0	1	30,336
Internet Cafe	0.208	0.406	0	1	30,340
Bank Branch	0.098	0.298	0	1	30,342

Note: Author's elaboration based on MINEDU's administrative records. The table reports the means, standard deviations, minimum and maximum values, and the sample size. Sample includes all maintenance civil servants who had not submitted their expense report at the beginning of the SMS campaign.

Table S6. Descriptive Statistics for External Validity experiment

Variables	Mean	Std. Dev.	Min	Max	N
Panel A: Pre-treatment Outcomes					
Compliance - August	0.669	0.369	0	1	1,113
Panel B: Outcomes					
Compliance - September	0.663	0.367	0	1	1,090
Compliance - October	0.660	0.394	0	1	1,075
Compliance - November	0.697	0.361	0	1	1,075
Compliance - December	0.730	0.376	0	1	1,075
Compliance - January	0.759	0.326	0	1	1,058
Panel C: Civil Servant Characteristics					
Sex (% Men)	0.432	0.496	0	1	1,116
Experience at CUNA MAS (years)	0.583	0.493	0	1	1,116
Postgraduate	0.192	0.394	0	1	1,116
Graduate	0.635	0.482	0	1	1,116
Technician	0.051	0.220	0	1	1,116
Another level of study	0.069	0.254	0	1	1,116
Language - Aimara	0.010	0.099	0	1	1,116
Language - Spanish	0.665	0.472	0	1	1,116
Amazonian language	0.004	0.067	0	1	1,116
Language -Quechua	0.264	0.441	0	1	1,116
Identified - Province	0.279	0.449	0	1	1,116
Identified - Coast	0.036	0.186	0	1	1,116
Identified - Jungle	0.111	0.314	0	1	1,116
Identified -Sierra	0.518	0.500	0	1	1,116
Not Peruvian	0.004	0.060	0	1	1,116
Panel D: Living Conditions					
Stereo	0.675	0.469	0	1	1,116
Television	0.874	0.332	0	1	1,116
Computer	0.748	0.434	0	1	1,116
Washing machine	0.252	0.434	0	1	1,116
Bicycle	0.211	0.409	0	1	1,116
Panel E: CUNA MAS Program					
How many CS record the information?	7.1	2.6	1	10	1,057
Has SAF delivered you a tablet?	0.905	0.293	0	1	1,116
Functional tablet	0.877	0.328	0	1	1,116

Note: Author's elaboration based on CUNA MAS's administrative records and a survey designed for this study. The table reports the means, standard deviations, minimum and maximum values, and the sample size.

Table S7. Randomization Balance Analysis for Follow-Up Experiment 2016

Variables	Control	Descriptive Social Norm	Injunctive Social Norm	Social Benefit	Joint Hypothesis
Panel A: Pre-treatment Outcomes					
Submitted to Maintenance Committee at Week 15	0.888	0.885	0.877	0.883	0.271
	(0.004)	(0.003)	(0.004)	(0.003)	
Submitted to Oversight Committee at Week 15	0.887	0.884	0.875	0.882	0.215
	(0.004)	(0.003)	(0.004)	(0.003)	
Submitted Technical Form at Week 15	0.769	0.772	0.764	0.763	0.467
	(0.006)	(0.004)	(0.006)	(0.005)	
Submitted Commitment Act at Week 15	0.743	0.748	0.739	0.735	0.178
	(0.006)	(0.004)	(0.006)	(0.005)	
Submitted Expense Report at Week 15	0.107	0.109	0.107	0.107	0.966
	(0.004)	(0.003)	(0.004)	(0.003)	
Submitted Oversight Report to the Week 15	0.005	0.005	0.005	0.007	0.468
	(0.001)	(0.001)	(0.001)	(0.001)	
With Approved Expense Report at Week 15	0.011	0.013	0.013	0.015	0.253
	(0.001)	(0.001)	(0.001)	(0.001)	
Bank Balance at 14/05	2670.5	2725.2	2619.5	2704.1	0.406
	(54.644)	(37.522)	(51.839)	(43.763)	
Bank Balance at 11/06	1556.3	1568.7	1539.4	1541.6	0.917
	(43.494)	(29.240)	(41.168)	(34.386)	
Panel B: Maintenance CS Characteristics					
Sex (% Men)	0.465	0.465	0.454	0.466	0.481
	(0.007)	(0.005)	(0.007)	(0.005)	
Age	44.1	43.8	44.0	44.1	0.186
	(0.169)	(0.117)	(0.160)	(0.132)	
Allocation Transfer	6973.8	6976.5	6974.1	6997.9	0.972
	(51.716)	(34.602)	(49.112)	(40.010)	
Panel C: School Characteristics					
Classrooms	4.9	4.9	4.9	4.9	0.756
	(0.077)	(0.050)	(0.071)	(0.058)	
Students	79.4	83.4	94.7	84.6	0.516
	(2.537)	(2.678)	(14.341)	(5.269)	
Bathroom Connected to Public Drainage System	0.345	0.338	0.34	0.342	0.793
	(0.007)	(0.004)	(0.006)	(0.005)	
Bathroom Connected to Septic Tank	0.232	0.241	0.248	0.242	0.274
	(0.006)	(0.004)	(0.006)	(0.005)	
Bathroom Connected to a Black Well	0.272	0.272	0.272	0.267	0.881
	(0.006)	(0.004)	(0.006)	(0.005)	
Bathroom Connected to River, Ditch or Canal	0.031	0.03	0.027	0.029	0.654
	(0.002)	(0.002)	(0.002)	(0.002)	
No Bathroom	0.069	0.07	0.065	0.068	0.594
	(0.003)	(0.002)	(0.003)	(0.003)	
Total Land Area	6133.0	6427.8	6214.4	7099.4	0.699
	(493.965)	(342.430)	(494.118)	(923.863)	
Fully Fenced	0.303	0.298	0.296	0.287	0.196
	(0.006)	(0.004)	(0.006)	(0.005)	
Partially Fenced	0.246	0.233	0.246	0.249	0.033
	(0.006)	(0.004)	(0.006)	(0.005)	
Unfenced	0.4	0.419	0.411	0.411	0.137
	(0.007)	(0.005)	(0.006)	(0.005)	
Number Educ-Admin Spaces	7.2	7.1	7.1	7.1	0.936
	(0.116)	(0.076)	(0.112)	(0.089)	
Number of Buildings	2.0	2.0	2.0	2.1	0.588
	(0.029)	(0.019)	(0.029)	(0.023)	
Average Leaks in Pavilions	1.1	1.1	1.1	1.1	0.827
	(0.021)	(0.013)	(0.018)	(0.016)	
Average Leaks	1.0	1.0	1.0	1.0	0.32
	(0.018)	(0.012)	(0.017)	(0.015)	

Table S7. Randomization Balance Analysis for Follow-Up Experiment 2016 (Continued)

Variables	Control	Descriptive Social Norm	Injunctive Social Norm	Social Benefit	Joint Hypothesis
Panel D: District Characteristics					
Altitude	2053.3	2039.9	2035.6	2050.2	0.883
	(20.100)	(13.411)	(19.047)	(15.589)	
Area (% Rural)	0.354	0.347	0.344	0.349	0.734
	(0.007)	(0.004)	(0.006)	(0.005)	
Electricity	0.812	0.806	0.807	0.814	0.501
	(0.005)	(0.004)	(0.005)	(0.004)	
Public Drinking Water Network	0.640	0.643	0.646	0.653	0.408
	(0.007)	(0.005)	(0.006)	(0.005)	
Public Drainage Network	0.370	0.363	0.363	0.369	0.706
	(0.007)	(0.005)	(0.006)	(0.005)	
Internet Cafe	0.203	0.209	0.208	0.210	0.791
	(0.006)	(0.004)	(0.005)	(0.004)	
Bank Branch	0.099	0.101	0.093	0.097	0.411
	(0.004)	(0.003)	(0.004)	(0.003)	
Observations	5,325	11,833	5,916	8,873	

Note: Authors' elaboration. Sample includes all maintenance civil servants who had not submitted their expense report at the beginning of the SMS campaign. For each treatment arm, means and standard errors are reported for each pre-treatment variable. Final column is the p-value for the test of equality of means across all groups. Table S1 in the Online Appendix contains the variables' full definitions.

Table S8. Randomization Balance Analysis for External Validity Experiment

	(1)	(2)	(3)	(4)
Variable	**Control**	**Social Norm**	**Monitoring**	**Joint Hypothesis**
Panel A: Pre-treatment Outcomes				
Compliance - August	0.683	0.664	0.657	0.739
	(0.024)	(0.024)	(0.025)	
Panel B: Civil Servant Characteristics				
Sex (% Men)	0.410	0.433	0.453	0.688
	(0.035)	(0.035)	(0.035)	
Experience at CUNA MAS (years)	0.595	0.576	0.579	0.916
	(0.035)	(0.035)	(0.034)	
Postgraduate	0.195	0.219	0.163	0.318
	(0.028)	(0.029)	(0.025)	
Graduate	0.608	0.652	0.648	0.598
	(0.034)	(0.033)	(0.033)	
Technician	0.062	0.028	0.061	0.100
	(0.016)	(0.011)	(0.016)	
Another level of study	0.065	0.076	0.067	0.902
	(0.017)	(0.019)	(0.018)	
Language - Aimara	0.010	0.008	0.011	0.968
	(0.008)	(0.006)	(0.008)	
Language - Spanish	0.639	0.697	0.661	0.467
	(0.034)	(0.033)	(0.034)	
Amazonian language	0.005	0.003	0.005	0.841
	(0.004)	(0.003)	(0.005)	
Language -Quechua	0.270	0.261	0.261	0.974
	(0.031)	(0.032)	(0.031)	
Identified - Province	0.252	0.287	0.299	0.541
	(0.030)	(0.032)	(0.032)	
Identified - Coast	0.026	0.039	0.043	0.562
	(0.010)	(0.014)	(0.014)	
Identified - Jungle	0.094	0.129	0.112	0.480
	(0.019)	(0.023)	(0.021)	
Identified -Sierra	0.553	0.514	0.485	0.387
	(0.035)	(0.035)	(0.035)	
Not Peruvian	0.005	0.006	0.000	0.367
	(0.005)	(0.006)	(0.000)	
Panel C: Living Conditions				
Stereo	0.691	0.646	0.685	0.574
	(0.032)	(0.033)	(0.033)	
Television	0.857	0.893	0.872	0.527
	(0.024)	(0.021)	(0.024)	
Computer	0.706	0.764	0.776	0.237
	(0.032)	(0.030)	(0.030)	
Washing machine	0.255	0.264	0.237	0.811
	(0.031)	(0.030)	(0.029)	
Bicycle	0.216	0.222	0.197	0.814
	(0.030)	(0.029)	(0.027)	
Panel D: CUNA MAS Program				
How many CS record the information?	7.316	7.000	7.011	0.390
	(0.183)	(0.182)	(0.193)	
Has SAF delivered you a tablet?	0.875	0.930	0.912	0.196
	(0.024)	(0.019)	(0.020)	
Functional tablet	0.834	0.919	0.883	0.034
	(0.026)	(0.020)	(0.022)	
Observations	385	356	375	

Note: Authors' elaboration. For each treatment arm, means and standard errors are reported for each pre-treatment variable. Final column is the p-value for the test of equality of means across all groups.

Table S9. Lee (2009) Bounds-Withdrawal of Maintenance Funds in the Benchmark Experiment

	(1)	(2)	(3)	(4)	(5)
Variable	**Withdraw Something**	**Withdraw 50%**	**Withdraw 80%**	**Withdraw 95%**	**Withdraw 99%**
SMS	0.043	0.534	0.916*	1.048**	1.458**
	(0.081)	(0.463)	(0.485)	(0.523)	(0.572)
Lower	0.043	0.531	0.911**	1.037**	1.463***
	(0.083)	(0.400)	(0.432)	(0.460)	(0.478)
Upper	0.092	0.579	0.960**	1.085**	1.512***
	(0.091)	(0.402)	(0.434)	(0.461)	(0.479)
Control mean	99.693	92.546	91.097	89.783	88.743
Controls	No	No	No	No	No
Observations	21.012	21.012	21.012	21.012	21.012

Note: Bounds are for outcomes indicated in each column and give Lee bounds under extreme assumptions about excess attrition in the National Bank data. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% respectively.